# USING THE MULTILEVEL GENERALIZED MIXED MODEL TO IMPUTE MISSING ACCELERMOMETRY

Rebbecca P. Wilson

A dissertation submitted to the faculty at the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Public Health in the Department of Biostatistics in the Gillings School of Global Public Health.

Chapel Hill
2018

Approved by:

Shrikant Bangdiwala

Suchindran Chirayath

Kelly R. Evenson

John Preisser

Daniela Sotres-Alvarez

## ABSTRACT

Rebbecca P. Wilson:
Using the Multilevel Generalized Mixed Model to Impute Missing Accelerometry
(Under the direction of Shrikant Bangdiwala and Daniela Sotres-Alvarez)

Accelerometers provide objective measures of physical activity and sedentary behavior. Typically, the device is worn for one week during all waking hours to measure physical activity counts for a period time (e.g., minute). A challenge is accounting for time when the accelerometer is not worn which can bias assessments of physical activity and sedentary behavior. To circumvent this, researchers will limit analysis to participants with a minimum number of adherent days with sufficient wear time and average these days. Excluding accelerometer nonwear assumes missing completely at random (MCAR); yet, sedentary behavior and physical activity are related to nonwear.

We proposed imputing counts/min for nonwear using a multilevel generalized mixed model (MGMM) and account for multivariate counts under a complex survey design. Using data from the Hispanic Community Health Study/ Study of Latinos (2008 – 2011), and the HCHS/SOL Youth Study (2012 – 2014), we utilize these methods to: (1) compare accelerometer wear and nonwear data in intervals of the day, (2) determine an association between accelerometer average counts/min and BMI, and (3) evaluate the different models using percent relative bias in simulated data.

Our results showed that (1) accelerometer average counts/min were higher for wear versus nonwear segments in an interval, thus, we concluded that the MCAR assumption of the ad hoc approach was not tenable. (2) The MGMM indicated a clear association between average count/min and BMI when missing values were imputed at the interval level. (3) The percent relative bias did not show enough evidence to support a smaller value for MGMM imputation evaluation models that were concordant with MGMM generated data. We concluded that imputing missing values at

the smallest unit possible (e.g., interval), and then aggregating at the participant level, may reduce the potential for making a type 2 error.

Further research in this area will greatly improve physical activity guidelines established using accelerometer data that better accounts for nonwear time.

## TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

## LIST OF ABBREVIATIONS

| | |
|---|---|
| ARBP | Additive regression, bootstrapping, and predictive mean matching |
| BLUP | Best linear unbiased prediction |
| BMI | Body mass index |
| CCA | Complete case analysis |
| CI | Confidence interval |
| d | Day(s) |
| e.g. | Exempli gratia |
| EM | Expectation-maximization |
| FCS | Fully conditional specification |
| GEE | Generalized estimation equations |
| GLLAMM | Generalized linear latent and mixed models |
| h | Hour(s) |
| HCHS/SOL | The Hispanic Community Health Study/Study of Latinos |
| i.e. | id est |
| IPW | Inverse probability weighting |
| kg | Kilogram(s) |
| m | Meter(s) |
| MAR | Missing at random |
| MCAR | Missing completely at random |
| MET | Metabolic equivalent of task |
| MGMM | Multilevel generalized mixed model |
| MI | Multiple imputation |
| ML | Maximum likelihood |

| | |
|---|---|
| MVPA | Moderate to vigorous physical activity |
| NB | Negative Binomial |
| NHANES | National Health and Nutrition Examination Survey |
| NMAR | Not missing at random |
| PSU | Primary sampling unit |
| RE | Relative efficiency |
| s | Second(s) |
| SD | Standard deviation |
| SE | Standard error |
| yr | Year(s) |
| ZIPLIN | Zero-inflated Poisson and log-normal |

# CHAPTER 1: INTRODUCTION

## 1.1 Background

A physically active lifestyle among children and adults promotes short- and long-term physical and mental well-being. Physically active individuals are healthier than those who are habitually inactive. They are less likely to become obese and develop related chronic illnesses such as type 2 diabetes, cardiovascular disease, depression, and some cancers.[1–3] Physical activity is generally described as any body movement that works muscles and increases energy expenditure beyond what the body expends at rest.[4] Any waking activity predominately done while in a sitting or reclining position while awake "that expends energy at or minimally above a person's resting level" is characteristic of sedentary behavior.[4] Self-reported physical activity questionnaires and a monitoring device, known as an accelerometer, are two of the most common methods for measuring physical activity in large epidemiologic studies. These device-based and reported methods of assessment both have strengths and weaknesses in different research contexts.[5]

One strength of self-report instruments is that they are a cost-effective way to gather physical activity data on thousands or hundreds of thousands of individuals.[6] An additional strength of self-report instruments is that they provide an integrated proxy measure for bodily motion that may incorporate elements of psychosocial and environmental context, activity purpose, perceived time-use and intensity of effort.[7] One limitation of self-reports of physical activity is that they are less reliable for activities of light-intensity, which tend to be poorly reported.[8] Accelerometer strengths include the accuracy with which the device is able to capture physical activity and the ability to record large amounts of data. One limitation of hip-worn devices is that they do not capture upper-body movement well.[6] They primarily measure locomotor activity.[6] Lastly, accurately and precisely summarizing accelerometer data in the presence of missing data creates a variety of analytical challenges.

Accelerometers are devices that can be used to estimate intensity, frequency, and duration of physical activity and frequency and duration of sedentary behavior. Typical accelerometer measurement protocols require participants to wear the device on the hip for several consecutive days, during waking hours, in order to obtain reliable estimates of physical activity and sedentary behavior, as habits can vary drastically throughout a day and week.[9] Devices, such as the Phillips Respironics Actical accelerometer, worn on the hip, can be used to estimate physical activity and sedentary behavior. These devices measure acceleration in units of gravity ($g \approx 9.81\ m/s^2$) at a specific frequency (e.g., 30 times/s or 30 Hz).[10] The company aggregates accelerations to activity counts per user-specified unit of time, known as an epoch.[10] The analysis of count data presents a number of challenges which are described in the next section.

### 1.2 Accelerometer Data and Research Challenges

A challenge in free-living accelerometer studies where the data is only available through a proprietary algorithm (e.g., Actical) is distinguishing sedentary behavior from nonwear since theoretically, both can register zero counts/epoch (e.g., 0 counts/60 s epoch).[7] Counts have historically been the dominant unit of measure of accelerometer activity.[7] Raw signal data from three axes is commonly processed using manufacturer proprietary algorithms to calculate counts to facilitate the use of their devices.[7] Inconsistencies in recorded counts/min among different accelerometers have been reported.[7,11] Thus, the challenge of distinguishing sedentary behavior from nonwear is only applicable to certain accelerometers that apply a proprietary conversion algorithm to the raw data. The Actical, the accelerometer used for this study, is one such device. Nonwear time, resulting in a period of consecutive zero counts, later becomes missing data. This subsequent missing data is a result of data aggregation methods employed by investigators to reduce the bias that may arise when all observed data is used to compute summary statistics. For example, an estimate for total counts for a particular day that is based on all observed data will underestimate the true level of physical activity if the device is worn for only part of the day.[9] To circumvent this potential bias, researchers use the inclusion criteria that participants have a minimum number of days each with a minimum number of hours of wear (e.g., ≥ 4 d with ≥ 8 h wear time/d) in order to be included in the final analysis sample. Participants with accelerometer data that do not meet the aforementioned criteria are commonly categorized as *missing* and/or non-adherent, and subsequently excluded from analyses.[9,12]

Common practice first summarizes accelerometer data at the day level (e.g., sedentary min/d), and then averages across days only for *non-missing* days.[13,14] The concern with this approach is that it assumes physical activity and sedentary behavior are missing completely at random (MCAR) during nonwear time; however, sedentary behavior and physical activity may be related to nonwear. Missing completely at random is one type of missing data mechanism. It means that the underlying cause of the missing data is unrelated to the value of any variables, observed or missing. Failing to account for the proportion of missing data, pattern, and mechanism can lead to potentially biased parameter estimates and weaken the generalizability of results.[15] In general, if the missing data rate is greater than 5-10%, not accounting for missing data is not advised.[16] Substituting missing values with plausible values that minimize bias, maximize the use of available information, and yield good estimates of uncertainty is ideal.[17]

The classification of participants as adherent versus non-adherent is the most common (i.e., ad hoc) approach for analyzing accelerometer data according to the literature; directly addressing missing data via an imputation method is less conventional and experimental.[9,12,14,18,19,20] The prominent components that should be addressed as part of a protocol for analyzing accelerometer data, as presented in the literature, can be grouped into the following six categories:

1) Bias reduction
2) Use of all available information maximized
3) Accounts for autocorrelation, between-, and within-participant variation
4) Accounts for complex survey design
5) Accessible to a wide range of researchers
6) Expands understanding of nonwear time

**Bias**

Bias is an important criterion for assessing the performance of an imputation technique. Bias is the average difference between the estimator and the true values (e.g., over- or underestimation of parameter estimates).[9] Smaller values indicate less bias. An unbiased estimator is not always the best-case scenario if it also results in greater variability (i.e., decreased precision). Several prior imputation methods were successful in reducing bias, in addition to increasing precision when applied to accelerometry.[9,19,21,22] Catellier experimented with single value imputation, using the expectation-maximization (EM) algorithm, and multiple imputation (MI) of metabolic equivalents for minutes in moderate-to-vigorous physical activity (MVPA) for different time periods (i.e., day, period of the day, weekday vs. weekend) in a physical activity intervention study of middle school girls in the U.S.[9] Using probability density estimates of missing values, EM preserves relationships with other variables. The E step of the algorithm uses other variables to impute a missing value. The M step checks to see whether the imputed value is most likely. If not, a new value is re-imputed that is more likely. This 2-step iterative process continues convergence (i.e., until the most likely value is reached).

Multiple imputation MI is a form of imputation that reflects uncertainty about an imputed value by replacing each missing value with $m > 1$ plausible values. The final result is m plausible values, versus a single plausible value, where missing data are replaced by independent random draws from the predictive distribution of the data matrix given the observed data. The m completed data sets (m data matrices with observed and imputed values) are analyzed separately using standard complete-data methods, and the results combined in a manner that takes the imputation variability into account. Both methods for imputation, when applied to simulated MCAR and not missing at random (NMAR) data, were affected by the proportion of missing data, correlation of MVPA across days of the week, and the missing data mechanism.[9] Not missing at random data is another type of missing mechanism. It means that the underlying cause of the missing data depends on the missing value itself and possible observed values as well.[23]

Similar to Catellier's findings, Lee (2013) found that his combined imputation approach performed better with fewer missing data points, shorter missing time periods, and higher correlation between activity levels on the same day.[19] Lee (2013) performed MI via additive regression, bootstrapping, and predictive mean matching (ARBP) of missing accelerometer data. Additive regression is a semiparametric modeling approach. Nonlinear, optimal transformations of both the predictors and the response to maximize correlation between the transformed response and the sum of the transformed predictors. Bootstrapping is a statistical technique that involves sampling with replacement. The ARBP imputation approach involves 4 steps: (1) for each variable with a missing value (target variable) draw a sample with replacement from the observations in the entire dataset; (2) fit an additive model to predict the target variable; (3) use the fitted model to predict the target variable in all the original observations; and (4) impute each missing value of the target variable with the observed value whose predicted transformed value is closest to the predicted transformed value of the missing value (predictive mean matching).[18] Lee's approach produced unbiased estimates of total volume of activity per day when counts/min were imputed at the day level. However, Liu, using the same data source and imputation method as Lee (2013), for accelerometer steps at the epoch level, was not able to successfully reduce bias.

Liu used the ARBP approach to impute accelerometer steps/min at the 60-s epoch level. Based on measures of relative bias, the ARBP imputation approach tended to slightly underestimate the mean, standard deviation, and median with a large degree of underestimation for the between person variance.[18] Imputing extreme steps was one limitation for this approach.[18] The objective for Liu's method was to choose an imputation method that minimized model-based assumptions, however, the high dimensionality of epoch-level data and failing to accounts for participant-specific effects may have adversely affected bias. Morris was successful in statistically characterizing the behavior of metabolic equivalents (METs), a measure of energy expenditure, from the triaxial TriTrac-R3D accelerometer at the epoch level, however the lack of clarity on bias makes it challenging to assess the performance of his wavelet-based functional mixed model (WBFF) method.[24]

A wavelet is a wave-like oscillatory function used for signal processing. Wavelets can be used to perform nonparametric regression to denoise (restore a signal which has been corrupted by some random process) a function of interest.[25] Accelerometer data recorded at the epoch level over time is an example of functional data. Functional data have a high frequency of measurements (possibly in multiple dimensions), and are smooth, complex processes. Functional mixed models consist of random functions; fixed effects functions and random effect functions. A Bayesian (estimation technique) model using the complete data was fit. Model parameter estimates were generated using complete data to impute missing values using Bayesian estimation. Posterior samples were used to estimate model parameters.[24] Morris was successful in maximizing the use of all available information.

**Use of all available information**

In general, sampling variability decreases as the sample size increases thus it is ideal to avoid discarding any data. Maximizing the use of all available information is a desirable characteristic of any imputation procedure. This implies days without a minimum number of hours of wear time be included in analyses. In addition to Morris, Lee (2013), and Lee (2016) maximized the use of all available information. Lee (2013) developed a combined approach to impute missing data from an adherent day by using both the information from adherent days and available data from days classified as non-adherent.[19]

Lee (2016) also used accelerometer counts from NHANES (2003-2004) to demonstrate a novel imputation method. The Zero-inflated Poisson and Lognormal (ZIPLN) model framework was used to specify the predictive distribution from which plausible missing values were drawn. Imputation was performed using the fully conditional specification (FCS). With FCS, regression models are specified for each variable with missing values, conditional on all other variables in the imputation model.[26] In addition to maximizing the use of all available information, accounting for the different sources of variability during imputation can help improve estimates of missing values.

**Accounts for autocorrelation**

Accounting for the different sources of variability during the imputation process can help improve the reliability of estimates of missing values. The methods proposed by Morris and Xu account for the different sources of variability via random effects.[21,24] Lee (2016) added an autoregressive term (i.e., lag variable) to the Poisson mean portion of the ZIPLN model.[22] In contrast, even though Catellier, Lee (2013), and Liu used repeated measures data, there was no direct account for within- and between-participant variability. Xu included a random intercept and random slope for each participant in the linear mixed model to impute missing counts per day from a 2016 weight loss intervention trial for overweight postmenopausal breast cancer survivors.[21] The rationale for using random effects approach is to use information from repeated measures within a participant as well as from across other participants' daily activity profiles.[21]

### Accounts for complex survey design

Accounting for the complex survey design during imputation is not applicable to or necessary for every accelerometer study but may be of interest to some researchers. Three of the reviewed studies, Lee (2013), Lee (2016), and Liu used accelerometer data from NHANES and only Liu incorporated the sample weight and primary sampling units (PSU) in the imputation model.

### Accessibility

Methods that are accessible to a wide range of researchers can be a challenging balance to attain. Moreover, the concept of accessibility is subjective. All of the studies reviewed are fairly straightforward to understand and implement with the exception of Morris'. Each study provides ample online supplemental documentation as well as SAS or R scripts that can be easily implemented. The sophistication of Morris' method does not appear to be a feasible endeavor as it is computationally and memory intensive and requires at least an intermediate level of understanding of functional data analysis and wavelets.

### Improves the understanding of nonwear time

A method that expands our understanding of nonwear time might help better characterize what may be happening during those periods. The traditional approach of excluding participants that are categorized as non-adherent assumes that the participants with missing data can be thought of as a random selection of participant that are categorized as adherent. None of the studies reviewed address the possible differences in the accumulation of accelerometer counts during wear and nonwear periods and during the imputation process.

The aforementioned studies do not provide guidelines for how to comprehensively address missing accelerometer data topics 1-6. There is a lot of discordance in the literature. The traditional approach of excluding participants that do not meet the wear time inclusion criteria might overestimate the true level of activity and distort patterns of behavior in the population. This could be problematic in research studies. Depending on the variable of interest, inaccurately describing a particular population based on biased physical activity and sedentary behavior estimates could lead to misinformed and detrimental decision making. For example, inaccurately categorizing a population as having met some physical activity guideline, as opposed to not having met it and thus requiring intervention, could lead to members of the population not receiving an intervention. This dissertation will address the components, items 1-6, that should be included as part of a protocol for analyzing accelerometer data in three separate papers.

The purpose of Paper 1: "A Repeated Measures Imputation Approach for Missing Accelerometer Data" is to provide an accessible statistical method to impute missing accelerometer data that incorporates the use of all available data, variability from within and between participants, and is well suited to account for multivariate count data. Treating counts/min as a rate and assuming counts are Poisson distributed, the innovation of this work is that it improves our understanding of how total counts, counts/min, and sedentary behavior, per day, may be accumulated during nonwear versus wear time by modeling the offset as a covariate.

The purpose of Paper 2: "A Simulation Study to Assess the Performance of the Multilevel Generalized Mixed Model Method used to Impute Accelerometer Missing Data in the HCHS/SOL Study" is to assess the performance of the proposed method using a measure of percent relative bias. Smaller values of this metric are indicative of greater accuracy.

The purpose of Paper 3: "Application of the Multilevel Generalized Mixed Model and other approaches to Address Accelerometer Missing Data: The HCHS/SOL Youth Study" is to compare the multilevel generalized mixed model imputation method with five different methods used for handling missing accelerometer data. We will also investigate whether or not imputing data at the interval of the day, day, or participant level makes a difference on point estimates and tests for associations between average counts/min and body mass index (BMI) (kg/m$^2$).

### 1.3 Missing Data

This section provides a technical overview of missing data in the longitudinal/repeated measures setting. The accelerometer data used for this study fit the repeated measures framework as each participant had a possible maximum of six days of data recorded at the 60-s epoch level. Moreover, the data were aggregated at different levels for analysis (e.g., counts/min/interval), maintaining the repeated measures structure. Missing data are ubiquitous in many studies that rely on data collection as part of the research process, regardless of discipline. The term 'missing data' is generally used to indicate that an intended measurement could not be obtained.[23] approaches for handling missing data range from simple to sophisticated. Before we can expound upon common strategies employed to address missing data, a discussion of missing data mechanisms, patterns, and quantity is necessary to understand how associated assumptions affect the use of certain procedures. Much of the following notation and methodology is borrowed from Fitzmaurice, Laird, and Ware's *Applied Longitudinal Data Analysis* unless otherwise stated.[23]

Considering a longitudinal data framework, assume the set of possible responses for a particular subject *i* can be represented by the $n \times 1$vector denoted by

$$Y_i = (Y_{i1}, Y_{i2}, \dots, Y_{in})';$$

Due to missing data, some of the components of $Y_i$ may not be observed for some participants which can be further denoted by an $n \times 1$vector

$$R_i = (R_{i1}, R_{i2}, \dots, R_{in})';$$

Where $R_{ij} = 1$ if $Y_{ij}$ is observed and $R_{ij} = 0$ if $Y_{ij}$ is missing. Additionally, there is an $n \times p$ matrix of covariates that are assumed to be fixed, $X_i$, and that correspond to $Y_i$ where any missing data in the covariates is not considered. The set of possible responses for each participant, $Y_i$, given

$R_i$, can be partitioned into two components $Y_i^O$ and $Y_i^M$ where $Y_i^O$ corresponds to observed responses and $Y_i^M$ corresponds to missing responses. The random vector $R_i$ can be thought of as a type of stratification variable that can be used to classify individuals in the population of interest into distinct subgroups based on missing data patterns. For example, suppose *n* = 3, the possible subgroups defined by $R_i$ are

$$(1,1,1)', (0,0,0)', (1,1,0)', (0,1,1)', (1,0,1)', (0,0,1)', (1,0,0)', (0,1,0)';$$

The missing data mechanism can be thought of as a probability model for the distribution of a set of response indicator variables (e.g., $P(R_{i3} = 1|Y_{i1}, Y_{i2}, Y_{i3}, X_i.)$.[23] The three types of missing data mechanisms are generally described in a hierarchical manner based on the relationship of $R_i$ and $Y_i$;

1. Missing Completely at Random (MCAR)
2. Missing at Random (MAR)
3. Not Missing at Random (NMAR)

### 1.3.1 Missing Completely at Random (MCAR)

Longitudinal data are assumed to be missing completely at random (MCAR) if the probability that $Y_{ij}$ is missing is independent of observed and unobserved components of $Y_i$, $Y_i^O, Y_i^M$. This can be expressed more formally as

$$P\left(R_i|Y_i^O, Y_i^M, X_i\right) = P(R_i|X_i);$$

An example with *n* = 3 is where covariate-dependent missingness is included in the definition of MCAR.

$$Y_i = (Y_{i1}, Y_{i2}, Y_{i3})'$$

$$R_i = (1,0,0)'.$$

If $Y_{i3}$ is MCAR, then

$$P(R_{i3} = 0|Y_{i1}, Y_{i2}, Y_{i3}, X_i) = P(R_{i3} = 0|X_i)$$

The probability that $Y_{i3}$ is missing does not depend on the observed value of $Y_{i1}$ or the values of $Y_{i2}$ or $Y_{i3}$ that, in theory, should have been obtained.[23] The key feature of MCAR is that the observed data can be thought of as a random sample of the complete data (possibly dependent upon covariates).[23] This implies that any analysis method that yields valid inferences in the absence of missing data will also yield valid inferences in distinct subpopulations with missing data patterns defined by $R_i$. Put another way, under the assumption of MCAR, the means, variances, covariances, and distribution of observed data do not differ from the corresponding moments or distribution for a subpopulation restricted to participants with fully-observed, complete data ("completers" or complete cases).

### 1.3.2 Missing at Random (MAR)

Next in the hierarchy of missing data mechanisms is the weaker, but still strong, assumption of missing at random (MAR). Longitudinal data are MAR if the probability that $Y_{ij}$ is missing depends on the set of observed components of $Y_i$, $Y_i^O$, but is conditionally unrelated to unobserved components of $Y_i, Y_i^M$. Data are MAR when $R_i$ is conditionally independent of $Y_i^M$ given $Y_i^O$;

$$P\left(R_i|Y_i^O, Y_i^M, X_i\right) = P\left(R_i|Y_i^O, X_i\right);$$

If $Y_{i3}$ is MAR, using the example of $n = 3$, then;

$$P(R_{i3} = 0|Y_{i1}, Y_{i2}, Y_{i3}, X_i) = P(R_{i3} = 0|Y_{i1}, Y_{i2}, X_i)$$

With MAR, observed data cannot be thought of as a random sample of the complete data. Due to the missing data mechanism's dependence on $Y_i^O$, the distribution of $Y_i$ in each of the distinct subpopulations defined by $R_i$ is not the same as the distribution of $Y_i$ in the group of "completers" or the target population. In other words, there may be systematic differences between observed, $Y_i^O$, and missing, $Y_i^M$, responses but these differences can be entirely explained by observed components of $Y_i^O$.

### 1.3.3 Not Missing at Random (NMAR)

The third, and most problematic, missing mechanism in the hierarchy of missing data mechanisms is not missing at random (NMAR). Longitudinal data are designated as NMAR when the probability that $Y_{ij}$ is missing depends on the missing value itself and possible observed values as well.[23] That is, the conditional distribution of $R_i$, given $Y_i^O$, is related to $Y_i^M$, and $P\left(R_i|Y_i^O, Y_i^M, X_i\right)$ depends on at least some components of $Y_i^M$.

For example, if $Y_{i3}$ is NMAR, using the example of *n* = 3, then;

$$P(R_{i3} = 0|Y_{i1}, Y_{i2}, Y_{i3}, X_i)$$

depends on the unobserved values of $Y_{i3}$.

### 1.3.4 Missing Data Patterns

Prospective cohort studies are typically designed to collect data on every participant at a planned sequence of occasions.[23] A missing data pattern refers to the order in which a set of observations that should have been collected is missing.[15,23] Suppose there are *n* response variables denoted by $Y_{i1}, Y_{i2}, Y_{i3}, \dots, Y_{in}$. If $Y_{ij}$ is missing and $Y_{ij+1}, Y_{ij+2}, Y_{ij+3}, \dots, Y_{in}$ are missing as well, the data are said to have a monotone missing data pattern.[15,23] This pattern is commonly observed in longitudinal studies when a participant drops out. When a participant drops out at a certain measurement occasion, all data for this participant will be missing at subsequent measurement occasions.[15,23] In contrast to the monotone missing data pattern is the non-monotone pattern which arises when data are missing intermittently.[23] Dealing with a monotone missing data pattern is computationally more flexible than with an intermittent missing data pattern.[15,27] Understanding the missing data pattern and underlying missing data mechanism can help inform which methods should be used to handle the missing data.

### 1.3.5 Amount of Missing Data

In addition to investigating the missingness mechanism and pattern in one's data, the amount of missing data should also be considered when determining how to analyze the data. Mishandling the missing values could lead to making inaccurate inferences about the data. If a case is missing any response or predictor variables it is simply excluded from analysis. This is the default approach for handling missing data in most statistical software packages. Analyzing only complete cases is generally regarded a statistically feasible approach as long as the number of incomplete cases is less than 5% of the sample.[28] This method also works well even when the number of incomplete cases exceeds 5% of the sample and the data are MCAR, however, this rarely occurs in practice. The alternative to analyzing only complete cases under ideal circumstances is to apply a statistical method to remediate the potential bias that arises when the assumption of MCAR is not tenable and the amount of missing data exceeds 5%.

Multiple imputation (MI), (Section 1.4.2), is often preferred to single value imputation (section 1.4.2) when the number of incomplete cases exceeds 5% of the sample and the assumption of MCAR is precarious.[28,29] The replacement of missing values with multiple plausible values ensures the uncertainty associated with the imputed values can be properly accounted for.[23] The general idea behind MI is to "fill in" or impute *m* plausible values for the missing observations thus creating *m* complete datasets that can be analyzed as if there were no missing data to begin with.[23] The number of imputations, *m*, needed for MI is a function of the rate of missing data.[15] The higher the rate of missing data, the more imputations needed. Relative efficiency (RE), in units of variance, can be used to determine the number of imputations needed.[15,29] Relative efficiency is approximately a function of the proportion of missing data, $\tau$, and the number of imputations, *m*.[30,29]

$$RE = \left(1 + \frac{\tau}{m}\right)^{-1}$$

### 1.3.6 Implications for Longitudinal/Repeated Measures Analysis

Valid inferences from analyses of incomplete data require assumptions about the missing data mechanism.[23] Failing to address the process that causes the missing data may lead to distorted results as individuals with missing values may systematically differ from those with completely observed data. In review of the previously discussed missing mechanisms; if data are MCAR, they are also MAR. Both mechanisms require that the missingness of a response be unrelated to the potentially missing value that, in principle, should have been obtained. The MAR assumption is less restrictive on $P(R_i)$ than MCAR and tends to be the more plausible assumption in studies with missing data.[23] However, the types of analyses that can be applied given the MAR assumption are more restrictive than those that can be applied given the MCAR assumption.[23] The types of analyses that can be applied under NMAR conditions are the most limited of the three mechanisms.

When the process that causes the missing data is unrelated to the missing values conditionally (MAR) or both conditionally and marginally (MCAR), the missing data mechanism can be ignored during the estimation process of interest. When the missing mechanism is *ignorable*, valid likelihood-based analyses can be obtained with no extra assumptions. The NMAR mechanism is referred to as *nonigorable* as it needs to be factored into the estimation process of interest. With regard to dependencies on observed data; under MAR conditions, missingness can be adequately predicted by the observed values in the data set, unlike missingness under NMAR conditions. Analyses conducted under the NMAR assumption that fail to model the missing mechanism are subject to bias.[30]

## 1.4 Techniques for Handling Missing Data

This section describes conventional methods for handling missing data.

### 1.4.1 Conventional Methods

**Listwise deletion**

Listwise deletion removes all data for a case from analysis if any value is missing. This is also known as an analysis of complete cases (CCA). With MCAR data, listwise deletion will decrease the analysis sample size however, bias should not be adversely affected.[15,17] A decrease in sample size typically leads to larger standard errors, wider confidence intervals, and a loss in power to test hypotheses. Estimated standard errors are reasonably accurate measures of the true standard errors.[17] If the data are MAR, listwise deletion may cause bias.[17]

**Pairwise deletion**

Also known as available cases analysis, pairwise deletion attempts to minimize the loss in information that occurs with listwise deletion. Pairwise deletion keeps as many cases as possible for each analysis.[17] Instead of removing cases with at least one missing value, pairwise deletion uses the available information for each case for each analysis. If data are MCAR, pairwise deletion should yield consistent and approximately unbiased parameter estimates.[17] If the data are MAR, pairwise deletion may produce biased estimates. Given the increased use of available information, pairwise deletion should typically be more efficient than listwise deletion but this may not always be the case.[15,17,29]

### 1.4.2 Advanced Methods

**Mean, median, or mode value imputation**

Mean, median, or mode value imputation is considered a straightforward approach to imputing missing data values based on observed data. This type of imputation will be explained using the mean but other point estimates could be applied in a similar manner. In the longitudinal data setting, missing values for an individual may be imputed by using their previously observed data or information can be borrowed from other participants with a similar outcome profile.[31] For example, in an accelerometer study with average counts per minute (per day) as the outcome, with a maximum of seven observation days; if a participant is missing three of the seven days, the mean from the observed four days may be substituted for each missing day. Using the same example, if a participant is missing six or all seven observations, mean values for each day may be imputed by using observed counts per minute values from participants with a similar profile as the one in question. Similar profiles could be determined by age, gender, and BMI for example.[31] Imputation using point estimates may be a reasonable approach when the correlations between measurement occasions is small and the overall percent of missing data is less than 10%.[32] The downside to using point estimates for imputation is that estimates of variance tend to be attenuated.[9] More broadly, single value imputation methods do not account for the uncertainty about the true value to impute which lead to an underestimation of the variance of the summary statistic.[9]

**Multiple Imputation (MI)**

Multiple imputation (MI), assuming MAR, appropriately reflects the sampling variability under a single model for non-response unlike mean value (or point estimate) imputation.[33] Multiple imputation refers to the procedure of replacing each missing value with $m$>1 plausible values which are randomly drawn from the posterior predictive distribution of the missing response values $Y_i^M$ given the observed data.[33] The next step in the MI procedure is to analyze each data set completed by imputation using standard complete-data methods.[9,33] Results are then combined taking into account the different sources of variability.[33] The variability associated with the combined estimate is comprised of the average within-imputation and between-imputation variance.[33] The only disadvantage of MI versus single value imputation, according to Rubin, is that MI requires more work to create the imputations and analyze the resulting data sets.[33] However, the benefits of MI far outweigh the costs considering the wide use and availability of statistical computing software.

**Inverse probability weighting (IPW)**

Multiple imputation methods replace missing values with $m$>1 plausible values randomly drawn from the conditional distribution of the missing data given the observed data.[23] An alternative approach to handling missing data, inverse probability weighting (IPW), does not require any assumptions about the conditional distribution of the missing data given the observed.[23] The bias that arises from analyzing complete cases only can be corrected by IPW when the missing at random (MAR) assumption holds. Complete cases are weighted by the inverse of their probability of being a complete case.[34] The probability of being a case is commonly calculated by fitting a logistic regression model to the observed data where the binary outcome variable equals one for being a case and it equals zero otherwise.[34] Covariates (and their relevant interactions) that help predict the probability of being a case should be included in the logistic regression model.[34] The IPW can then be applied to the standard fitting procedures.[23,34]

**Expectation-maximization (EM)**

The EM algorithm is a general iterative algorithm for maximum likelihood (ML) estimation for a broad range of analytical problems involving missing data. A brief discussion of ML estimation will be provided in order to facilitate our understanding of the EM method for missing data imputation. Suppose we have a random sample $G_1, G_2, G_3, \dots, G_n$ whose assumed probability distribution depends on an unknown parameter $\theta$. Our goal is to find a point estimator of $(G_1, G_2, G_3, \dots, G_n)$, such as the mean, where $\mu = \sum_{i=1}^{n} g_i \,/n$ such that $\mu$ is a good point estimate of $\theta$. A reasonable good estimate of unknown $\theta$ would be the value of $\theta$ that maximizes the probability or likelihood of getting the data we observed, $(g_1, g_2, g_3, \dots, g_n)$. Suppose the probability density function for each $G_i$ is $f(g_i, \theta)$, then the joint probability density function of $G_1, G_2, G_3, \dots, G_n$ is denoted $L(\theta)$. Where

$L(\theta) = P(G_1 = g_1, G_2 = g_2, \dots, G_n = g_n) = f(g_1, \theta) \times f(g_2, \theta) \times \cdots \times f(g_n, \theta) = \prod_{i=1}^{n} f(g_i, \theta)$ .

To find the value of $\theta$ that maximizes the likelihood, we differentiate $L(\theta)$ with respect to $\theta$. It is usually easier to differentiate the natural log of the likelihood function $\log L(\theta)$ with respect to $\theta$ than to take the derivative of $L(\theta)$. Now that we have some understanding of ML estimation, we will proceed with a description of the EM algorithm as it pertains to missing data imputation.

The EM algorithm for missing data imputation can be summarized by the following steps: (1) Replace missing values by estimated values, (2) estimate parameters, (3) re-estimate the missing values assuming the new parameter estimates are correct, (4) re-estimate parameters, and so forth, iterating until convergence.[33] The initial parameter estimate is calculated using the log-likelihood based on the complete data. Then the log-likelihood is estimated at each iteration of the algorithm (re-estimation) until the parameter estimates are stable and the log-likelihood value cannot be further improved (convergence).[33] The maximization, M-step, of the algorithm performs ML estimation of unknown parameter $\theta$ using the observed data. The expectation, E-step, of the algorithm finds the conditional expectation of the missing data given the observed data and current estimated parameters, and then substitutes these expectations for the missing data.[33]

## CHAPTER 2: A REPEATED MEASURES IMPUTATION APPROACH FOR MISSING ACCELEROMETER DATA

### 2.1 Introduction

A physically active lifestyle among children and adults promotes short- and long-term physical and mental well-being.[35] They are less likely to become obese and develop related chronic illnesses such as type 2 diabetes, cardiovascular disease, depression, and some cancers.[1,3,35,36] Self-report physical activity questionnaires and accelerometry are two of the most common methods for measuring physical activity in large epidemiologic studies. These device-based and reported methods of assessment both have strengths and weaknesses in different research contexts.[5] One strength of self-report instruments is that they are a cost-effective way to gather physical activity data on thousands or hundreds of thousands of individuals.[6] An additional strength of self-report instruments is that they provide an integrated proxy measure for bodily motion that may incorporate elements of psychosocial and environmental context, activity purpose, perceived time-use and intensity of effort.[7] One limitation of self-reports of physical activity is that they are less reliable for activities of light-intensity, which tend to be poorly reported.[8] Accelerometer strengths include the accuracy with which the device is able to capture physical activity and the ability to record large amounts of data. One limitation of hip-worn devices is that they do not capture upper-body movement well.[6] They primarily measure locomotor activity.[6] Lastly, accurately and precisely summarizing accelerometer data in the presence of missing data creates a variety of analytical challenges.

Typical accelerometer measurement protocols require participants to wear the device on the hip for several consecutive days, during waking hours, in order to obtain reliable estimates of physical activity and sedentary behavior, as habits can vary drastically throughout a day and week.[37] Devices, such as the Phillips Respironics Actical accelerometer, worn on the hip, can be used to estimate physical activity and sedentary behavior. These devices measure acceleration in units of gravity ($g \approx 9.81$ m/s$^2$) at a specific frequency (e.g., 30 times/s or 30 Hz).[10] The company aggregates accelerations to activity counts per user-specified unit of time, known as an epoch.[10] The analysis of count data presents a number of challenges which are described in the next section.

A challenge in free-living accelerometer studies where the data is only available through a proprietary algorithm (e.g., Actical) is distinguishing sedentary behavior from nonwear since theoretically, both can register zero counts/epoch (e.g., 0 counts/60 s epoch).[7] One classification of sedentary behavior is defined as 0 to 100 counts/min.[12] This classification is sensitive to the nonwear definition because a low or conservative threshold for identifying nonwear time is more likely to classify sedentary periods as nonwear.[38,39] Nonwear time, resulting in a period of consecutive zero counts, later becomes missing data which can bias assessments of physical activity.[21] Estimates of physical activity are biased downward as counts are not recorded during nonwear.[9] To circumvent this potential bias, researchers sometimes analyze data only for participants with a minimum number of adherent (or valid) days, defined as having a sufficient amount of wear time in a given day (ad hoc approach).[9,12] Non-adherent (or invalid) days are labelled as "missing". Common practice first summarizes accelerometer data at the day level (e.g., counts/min/d), and then averages across days only for non-missing days.[13,14] The concern with this approach is that it assumes physical activity and sedentary behavior are missing completely at random (MCAR) during nonwear. In other words, if the MCAR assumption holds, point estimates and the distribution of observed physical activity data for excluded (i.e., non-adherent) days should not differ from the corresponding point estimates or distribution for included (i.e., adherent) days. However, sedentary behavior and physical activity can be related to nonwear. For example, the MCAR assumption would be violated if participants are more (or less) likely to wear the device during nonwear.[9] While directly addressing accelerometer missing data via some imputation method is less conventional, a number of researchers have had success in obtaining more reliable estimates of physical activity and sedentary behavior using various statistical methods.[9,18,19,21,22,24]

Several authors have proposed procedures to impute missing accelerometer data. Catellier et al. implemented a single value imputation method, the expectation-maximization algorithm, and multiple imputation (MI) to impute missing accelerometer data for the entire day and interval of the day.[9] This study showed that the performance of each imputation method depends on the proportion of missing data, the correlation of activity across days of the week, and the missing data mechanism.[9] Using all available data from adherent and non-adherent days, Lee and Liu performed MI of daily accelerometer counts/min and missing steps per 60-s epoch, respectively, using additive regression, bootstrapping, and predictive mean matching.[18,19] In addition to using data from adherent and non-adherent days, Xu proposed a mixed model technique for imputation, which accounts for variation between and within participants.[21] Some advanced techniques for imputing missing accelerometer data include Bayesian and zero-inflated models.[22,24]

An additional drawback of summarizing accelerometer data at the day or participant level only for participants with a sufficient amount of data is that variations in physical activity, sedentary behavior, and nonwear throughout a day are masked. Specifically, recent studies suggest that the impacts of physical activity and sedentary behavior, at the minute or hour level, or patterns across a week, on health indicators warrants further investigation.[1,2,40,41] In adult populations, there is growing evidence on the adverse effects of prolonged sedentary time as well as how it is accumulated throughout the day. Independent of total sedentary time and moderate-to-vigorous intensity activity, Healy et al. found that increased breaks in sedentary time were beneficially associated with waist circumference, triglycerides, and 2-h plasma glucose.[42] In healthy, young adults Altenburg et al.

demonstrated that interrupting prolonged sitting every hour may be important for acute health outcomes as well as reducing postprandial glucose and insulin levels.[43] Estimating physical activity and sedentary behavior, given missing data, is essential to understanding how the total volume and patterns of accumulation of physical activity and sedentary behavior affect health outcomes of interest.[40,44] Average counts/min is an indicator of average total volume of physical activity and was the main outcome of interest for this study.

The purpose of this study was to provide an accessible statistical method to impute missing accelerometer data that incorporates all the available accelerometry data, accounts for variability from within and between participants, and could account for multivariate count data under a complex survey design. We hypothesize that the rate of average counts/minute will be higher for wear segments and lower for nonwear segments within an interval. For this reason, we believe accelerometer wear time serves better as a predictor than an exposure variable (i.e., $\gamma \neq 1$).

This study is innovative in that it improves our understanding of how average counts/min may be accumulated during nonwear versus wear time. Modeling the exposure variable of wear time as the covariate log-wear time, and estimating its regression coefficient, $\gamma$, instead of assuming $\gamma$ =1, which is this default assumption for the coefficient of the offset in log-linear models for rates, in the multilevel generalized mixed model (MGMM) allows for greater flexibility to fit the data. Allowing $\gamma$ to equal something other than one implies that rate of average counts/min is not the same during wear versus nonwear segments within an interval. The proposed method of using model parameter estimates from the MGMM of counts/min based on wear time to predict missing counts/min values during nonwear time is appealing as it can be implemented in most statistical software packages.

### 2.2 Physical Activity Demonstration Example

#### 2.2.1 Study Population

The Hispanic Community Health Study/Study of Latinos (HCHS/SOL) is a multicenter community-based cohort study designed to examine the risk factors of chronic disease among Hispanics/Latinos, aged 18 to 74, in the United States.[45] The HCHS/SOL enrolled 16,415 self-identified Hispanics/Latinos between March 2008 and June 2011 in four communities (the Bronx, New York; Chicago, Illinois; Miami, Florida; San Diego, California) in the United States. Recruitment was implemented through a two-stage area household probability sampling design.[45,46] Further details on the study design have been previously published.[45,46]

#### 2.2.2 Data Measures

Accelerometer data was obtained from participants at the four HCHS/SOL field centers. During the baseline clinic visit, participants were asked to wear the Actical™ accelerometer (model 198-0200-03, Minimiter Respironics®, Bend, OR) for seven days during waking hours. They were instructed to undertake usual activities while wearing the monitor on the hip and to remove it only for swimming, bathing, and sleeping. In addition to receiving instructions on proper wear during the clinic visit, participants were given written instructions and a phone number to call if questions arose during the 7-day monitoring period. At the end of the monitoring period, participants returned the Actical to the field center in person or via mail.

To standardize across sites, data recorded beginning at 5:00 a.m. on the day after the clinic visit, ending at a maximum of six days, were included. The length of day one was 19 hours (5:00 a.m.-11:59 p.m.) and the length of days two through six was 24 hours (12:00 a.m.-11:59 p.m.).[47]

The Actical accelerometer measures movement in all directions and was programmed to record counts and steps in 60-s epochs. Average counts/min was the primary measure for this study. Nonwear time was defined as at least 90 minutes of consecutive zero counts, with the allowance for intervals of up to two minutes of nonzero counts if the 30 minutes preceding and following these intervals were consecutive zero counts.[48]

### 2.2.3 Analytic Sample

A total of 16,415 participants were recruited for the original study. There were 12,750 participants classified as adherent (i.e., ≥10 h of wear time for ≥3 d). Of the adherent participants, 608 were excluded based on reporting having worked the night shift (after midnight) occasionally or regularly .[47] Moreover, only data recorded between 6:00 a.m. and 12:00 midnight each day (18 h) were included. Our interest was in estimating physical activity when participants have been instructed to wear the device (i.e., during waking hours). Thus, the start of our monitoring day at 6:00 a.m. Similar to Lee (2013), data from non-adherent days (i.e., <10 h/d) for adherent participants were included instead of excluding them as the ad hoc approach does. Due to the computational intensity of the method using the entire sample size, a stratified random sample of N = 1,000 participants was used for analysis.

### 2.2.4 Physical Activity Outcome

Each 18-h monitoring day was divided into six intervals of equal length: (1) 6:00 a.m.-9:00 a.m., (2) 9:00 a.m.-12:00 p.m., (3) 12:00 p.m.-3:00 p.m., (4) 3:00 p.m.-6:00 p.m., (5) 6:00 p.m.-9:00 p.m., and (6) 9:00 p.m.-12:00 midnight. The maximum possible wear time for each interval was three hours (180 min). Given our secondary interest in how average counts/min are accumulated throughout the day, we were interested in exploring intervals of the day as opposed to individual hours or minutes. We used six intervals of equal length for illustration purposes. Nevertheless, the length of the interval was arbitrary and could have been different. The physical activity outcome of interest was average counts/min in an interval, an indicator of average total volume of physical activity in an interval. This was calculated by taking the sum of all counts during wear time in the interval and dividing by the total wear time (in minutes) in that interval. Log-linear (Poisson) regression enables conversion of the dependent variable from a rate into a count. The log of the rate, average counts/min, is modeled as a function of a set of covariates in the multilevel generalized mixed model such that log(counts/min) = log(counts) – log(min) on the left-hand-side. The conversion of the dependent variable from a rate into a count is obtained by adding the log(min) term to the right-hand-side as an offset (or covariate) leaving log(counts) on the left-hand-side.

**Covariates**

Participant age and sex were self-reported during the baseline examination.[49] Participants' height was measured to the nearest centimeter and body weight to the nearest 0.1 kg. Body mass index was calculated as weight in kilograms divided by height in meters squared. Body mass index categories were defined as underweight (<18.5 kg/m$^2$), normal weight (18.5–24.9 kg/m$^2$), overweight (25.0–29.9 kg/m$^2$), and obese (≥30.0 kg/m$^2$).[49,50]

## 2.3 Methods

### 2.3.1 Overview of Statistical Analyses

The multilevel generalized mixed model method takes advantage of the richness in information provided by all available data among those with at least a certain number of adherent days. Under the assumption of missing at random (MAR), parameter estimates from the MGMM of counts per interval as a function of wear time, age (years), body mass index (BMI) (kg/m$^2$), sex, day (Sunday, Monday,…, Saturday), interval of the day, and strata were used to obtain unknown counts per interval during nonwear. Specifically, as the first step of the MI technique, the model parameter estimates were used to fill in five missing values in order to create five complete data sets. Counts per interval were assumed to be Poisson distributed. Age (years), BMI (kg/m$^2$), and wear time minutes were continuous; sex, day, interval, and strata were categorical. Intervals were nested within day and day within subject, hence the multilevel structure.

Given our interest in "filling in" missing values of accelerometer counts and the accumulation of counts during nonwear, we explored how to treat wear time in the imputation model. The options considered were (1) to specify log-wear time as an offset (which implies a coefficient equal to one) or (2) to estimate the regression coefficient by including log-wear time as a covariate. Our ultimate interest was in using the MGMM with wear time as a covariate. We have also included analyses for the complete cases (CCA) as defined by the HCHS/SOL (participants with ≥10 h of wear time for ≥3 d) and the marginal model using GEE for comparative purposes. The Z-statistic for the test of $H_0: \gamma = 1 \; vs. H_1: \gamma \neq 1$ under the MGMM and generalized estimating equations (GEE) model frameworks and corresponding 95% confidence intervals (CI) for $\gamma$ were used to guide our decision on how to include wear time in the imputation models. Using the model parameter estimates from the MGMM based on wear time, predicted values based on nonwear time were calculated then added to observed values to obtain imputed values for intervals that were partially observed or completely missing. "Completely missing" intervals were determined to be "missing" based on the nonwear definition used in the HCHS/SOL. Nonwear time for each interval was calculated as 180 minutes minus wear time. Counts for intervals that were completely observed were preserved (e.g., no imputation was performed). Under the assumption of MAR, the advantage of this approach over the ad hoc approach is that it utilized information from repeated measurements within an individual as well as borrowed information from other participants with the same covariate patterns to create more reliable estimates of physical activity.[21]

All statistical analyses were conducted using SAS software 9.4 (SAS Institute, Cary, NC) and Stata statistical software, Release 14 (StataCorp LP, College Station, TX). The generalized linear latent and mixed models (GLLAMM) procedure in Stata permits the user to incorporate unequal selection probabilities, in addition to fitting a hierarchical model, characteristic of many complex survey sampling designs in the form of scaled sampling weights $w_{ic}^*$.[51] The scaled weight, $w_{ic}^*$, was obtained by

$$w_{ic}^* = w_{ic}\frac{n_c}{\sum_i w_{ic}}$$

where $w_{ic}$ is the HCHS/SOL sampling weight for subject *i* in cluster *c* and $n_c$ is the number of participants within cluster *c*. Scaled weighted estimates and standard errors were compared with unweighted analyses. For the weighted analyses, imputation model parameter estimates were first obtained from the GLLAMM procedure using Stata and then exported to SAS to carry out the remaining imputation steps. The unweighted analyses were carried out in SAS.

### 2.3.2 Imputation Analysis Models

The following equation was used to model counts per interval as a function of the observed wear time and aforementioned covariates. The GEE model (not displayed) was identical to the MGMM with the exception of the random effects.

MGMM**:**

$$log\left(E\left(\frac{Y_{ijk}}{w_{ijk}^{\gamma}}\middle|\boldsymbol{b_i}\right)\right) = log\left(\frac{\mu_{ijk}}{w_{ijk}^{\gamma}}\right);$$

$$\begin{aligned} log(\mu_{ijk}) = \gamma log(w_{ijk}) + \beta_0 + \beta_1 Age_i + \beta_2 Sex_i + \beta_3 BMI_i \\ +\beta_4 I(Day_{ij} = Monday) + \cdots + \beta_9 I(Day_{ij} = Saturday) \\ +\beta_{10} I(Interval_{ijk} = 9am - 12pm) + \cdots + \beta_{14} I(Interval_{ijk} = 9pm - 12am) \\ +\beta_{15} I(Stratum_{ijk} = 1) + \cdots + \beta_{34} I(Stratum_{ijk} = 19) + b_{0i} + b_{1ij} \end{aligned} \quad (1)$$

Distributed as Poisson, let $Y_{ijk}$ be the observed accelerometer counts for participant *i*, on day *j*, in interval *k*, with the conditional mean of $Y_{ijk}$ given as $E(Y_{ijk}|\boldsymbol{b_i}) = \mu_{ijk}$ ; *i* = 1,2,3,…, 1,000, *j* = 1,…, 7, and *k* = 1,2…, 6, and where $\boldsymbol{b_i} = (b_{0i}, b_{1ij})$. Let $b_{0i} \sim N(0, \sigma_0^2)$ be the random intercept for participant *i* and $b_{1ij} \sim N(0, \sigma_1^2)$ be the random intercept for day *j* where interval *k* is nested within day *j* and day *j* is nested within participant *i*. The exposure, $w_{ijk}$, is the observed wear time minutes for each interval within each day for each participant. Interpret $\frac{\mu_{ijk}}{w_{ijk}}$ as "the expected counts per minute per interval given the random effects $\boldsymbol{b_i}$". The regression coefficient for the exposure is $\gamma$ and $\beta_0 - \beta_{34}$ are the regression coeffients for the remaining model covariates. There were 20 strata. Sex = female, day = Sunday, and interval = 6:00 a.m. – 9:00 a.m. were reference levels. The Z – statistic for the test of $H_0: \gamma = 1$ vs $H_1: \gamma \neq 1$ was calculated as $\frac{\hat{\gamma}-1}{se(\hat{\gamma})}$ where $\hat{\gamma}$, $se(\hat{\gamma})$, and the 95% CI for $\hat{\gamma}$ were obtained from the software output.

The following steps were undertaken:

**i.** Fit a multilevel generalized mixed model (MGMM) (or GEE) to wear data.

Fit the MGMM with observed counts, $Y_{ijk}$, as the dependent variable and log-wear time per interval, age (years), sex, BMI (kg/m$^2$), and day (Sunday, Monday,…, Saturday) as independent variables.

**ii.** Report the Z-statistic for $H_0: \gamma = 1\ vs. H_1: \gamma \neq 1$ along with the 95% CI for $\gamma$. If the 95% CI excludes the null value of 1 then include wear time in the imputation model.

**iii.** To obtain predicted counts for intervals that were entirely missing or partially observed, first substitute nonwear time, $w_{ijk}^{nw} = 180 - w_{ijk}$, for $w_{ijk}$ in equation (1). Random effects were estimated using best linear unbiased prediction (BLUP), also know empirical Bayes estimation. For intervals that were defined as nonwear and thus contained zero total counts, $w_{ijk}^{nw}$ = 180 minutes. Since participants were assumed to be wearing the accelerometer during the day, in accordance with the HCHS/SOL protocol, $w_{ijk}$ corresponding to the "missing" intervals were treated as wear time in the imputation model.

**iv.** Impute five values for entirely missing intervals where $Y_{ijk}^{imp}$ was imputed by random draw from the Poisson distribution with mean $\hat{\mu}_{ijk}$. In this scenario, $\hat{\mu}_{ijk}$ is the predicted value of mean counts given the random effects $\widehat{\boldsymbol{b}}_{\boldsymbol{i}}$ and $w_{ijk}$ = 180 minutes. The random Poisson component was generated five times for use as part of the MI step.

**v.** Impute five values for partially missing intervals where predicted counts, $Y_{ijk}^{mis}$, were calculated only for the missing portion of the interval. Similar to what was done in **iv**, $Y_{ijk}^{mis}$ was imputed by random draw from the Poisson distribution with mean $\hat{\mu}_{ijk}$. In this scenario, $W_{ijk} = 180 - w_{ijk}$. The final imputed value, $Y_{ijk}^{imp}$, is the sum of the observed counts, $Y_{ijk}$, and the predicted counts $Y_{ijk}^{mis}$.

### 2.3.3 Analysis with Complete Counts

We were interested in estimating the average counts/min per interval overall and by sex, age group, BMI group, interval, day of the week, and week days versus weekend days. We compared the results of the complete count analysis with those from imputed data. Choosing the imputation model covariates is an important initial step for any imputation procedure. To ensure robust analyses, the set of imputation model variables should be at least as large and comprehensive as the set of variables required for the analytic model.[28] Interactions and exponential forms of the variables should be considered and the analysis model should always be nested within the imputation model.[28] Failure to include an analysis model variable in the imputation model can lead to bias in the MI estimation and inference.[28] In this study, the main effects in the analysis model were the same as the main effects in the imputation model (minus the random effects).

**vi.** Analyze and summarize the data sets.

Complex survey regression was performed on the complete data sets with total counts per interval as the response using the method of GEE with a log link function. Complete data sets containing five imputations were formed by combining the imputed values from **v** with observed (non-imputed) counts from completely observed intervals. If a participant did not have at least some data on a day (i.e., 0 min wear time) then it was impossible to impute data for an interval in a day using the MGMM if there was no predicted value of the day intercept $b_{1ij}$. A complete data set contained counts for all six intervals for all days with a predicted value for $b_{1ij}$. The GEE was fit using the LOGLINK procedure in SUDAAN software, Release 11 (RTI International, Research Triangle Park, NC) with the robust variance estimator of Zeger and Liang (1986) in order to account for the complex survey design and to obtain valid standard errors for model parameter estimates.[53] After sorting the data sets by subject id, the strata and cluster variable (primary sampling unit ID) were used on the NEST statement. An exchangeable working correlation structure was estimated. Model parameter estimates from the complete data sets were combined and summarized using PROC MIANALYZE in SAS. The process of creating multiple data sets that have been "filled in" with plausible values for the missing data, applying an analysis model to the data sets, and combining the results, is broadly accepted as the optimal method for imputing MCAR and MAR data. The method of MI was established by Donald Rubin and has been well documented throughout the literature.[29,30,28,29] Relative efficiency (RE), in units of variance, was compared for five and ten imputations to decide on the optimal number of imputations. Relative efficiency is approximately a function of the proportion of missing data, $\tau$, and the number of imputations, *m*.[30,29]

$$RE = \left(1 + \frac{\tau}{m}\right)^{-1}$$

Results of the survey regression model using complete case analysis and data imputed using the MGMM and GEE with and without the offset as a covariate are presented.

### 2.4 Results

The analysis sample of n = 1,000 participants made up 7% of the overall HCHS/SOL population (Table 1). Approximately 50% of participants were women. Roughly 60% of this sample was between the ages of 25-54. Forty-one percent of the participants were classified as overweight and 38% as obese. Daily wear times were highly variable (range 10 to 18 h). Nonwear drastically decreased after 6:00am from 64% to 15% (Figure 1). The lowest levels of nonwear were observed from 12:00 p.m. – 7:00 p.m. Nonwear drastically increased after 8:00 p.m. Based on this information, participants were mostly compliant with the HCHS/SOL protocol with regard to when the monitor should be worn.

There was a total of 35,982 intervals (observations) for the 1,000 participants in the analysis sample. Of the 35,982 intervals, 4,230 (12%) were entirely missing. Four participants contributed five days and 996 participants contributed six days. Of the 35,982 total intervals, 21,208 (59%) were fully observed (i.e., no imputation needed), 4,230 (12%) were completely missing, and 10,544 (29%) were partially observed. Consistent with the observations of nonwear in Figure 1, the first interval had the most missing epochs (minutes) compared to the other intervals: (6am-9am: 22.3% n=1,342, 9am-12pm: 8.9% n=536, 12pm-3pm: 6.4% n=385, 3pm-6pm: 6.1% n=366, 6pm-9pm: 7.9% n=476, 9pm-12am: 18.7% n=1,123). Missing data drastically decreased between intervals 1 and 2, plateaued between intervals 2-5, and drastically increased between intervals 5 and 6 (Figure 2).

Figure 1 suggests nonwear for weekend days and week days was different between 6:00am-10:00am and 8:00pm-12:00 midnight. This can be explained by the variation of when participants wake up and go to bed during the week versus weekend. Examination of missing counts per interval data by day of the week revealed that Sunday and Saturday had the most missing data compared to weekdays (Sunday 17.4%, Saturday 12.1%, Monday 11.9%, Tuesday 11.0%, Wednesday 9.7%, Thursday 9.3%, Friday 9.8%). Thursday and Friday had the lowest amounts of missing data (Figure 3).

### 2.4.1 Imputation Model

Overall, there were no noticeable differences between the weighted (i.e., GLAMM) and unweighted analyses. Thus, results based on the unweighted analyses were reported. The imputation results based on the MGMM and GEE, for $\gamma = 1$ and $\gamma \neq 1$, were reported in Tables 2.3, 2.4, 2.8, and 2.9. In this study, the Poisson regression framework was used to model accelerometer counts per minute per interval where the exposure was accelerometer wear time minutes per interval. Due to the variability in accelerometer wear time across participants, days of the week, and intervals (range 0 – 180 min), Poisson regression enabled conversion of the dependent variable from a rate into a count. Under the assumption of $\gamma = 1$, the rate of counts/min per interval did not change within an interval, even when an interval contained both wear time and nonwear time. One of our suspicions at the outset was that this is not the case. We hypothesized the rate of counts/minute would be higher for wear segments and lower for nonwear segments within an interval. For this reason, we thought accelerometer wear time would serve better as a predictor than an exposure variable (i.e., $\gamma \neq 1$).

Information provided by the 95% CIs for $\gamma$ in the MGMM, [1.017, 1.150], and GEE, [1.136, 1.268], indicated that accelerometer wear time, log-wear time, should be included as a predictor in the imputation models. The Z-statistics from the imputation model parameter estimates for the MGMM and GEE for $H_0: \gamma = 1\ vs. H_1: \gamma \neq 1$ were 2.441 and 5.941 respectively (Tables 2.2, 2.5). The rate ratio of counts per interval was expected to increase by 8.3% for each unit increase in log-wear time while holding all other variables in the MGMM constant. The imputation model parameter estimates between the MGMM and GEE were fairly comparable with the exception of the intercepts. The estimates of intercepts for the MGMM, 5.792 for $\gamma = 1$ and 4.788 for $\gamma$ = 1.083, were smaller than the estimates for the GEE, 6.408 for $\gamma = 1$ and 5.402 for $\gamma$ = 1.202 (Tables 2.2, 2.5).

Estimates of mean counts/min per interval were robust to the number of imputations. There was no gain in using *m* = 10 versus *m* = 5 imputations based on relative efficiency (0.99 vs. 0.98) thus we only included results for *m* = 5 imputations. The MGMM resulted in lower imputed counts/min per interval compared to counts/min imputed using GEE and the complete case analysis (i.e., participants with ≥10 h of wear time for ≥3 d) (Table 2.3). The same trend was observed for total counts per interval (Table 2.9). This may be explained by the distinct target of inference associated with generalized linear mixed models and marginal models.[31] The regression coefficients in the GEE model describe the effects of the covariates on the population mean response while the regression coefficients in the MGMM describe how changes in a participant's mean response are related to within-participant changes in the covariates.[31] The average counts/min per interval calculated using the GEE are conditional only on the covariates and not on unobserved random effects. However, the mean counts/min per interval calculated using the MGMM are conditional on the random effects, which on average for incomplete intervals were negative. For the MGMM with $\gamma = 1.083$, the average of the random intercept predictions ($b_{0i} + b_{1ij}$) was lowest, -0.069, for the intervals that were entirely missing (i.e., fully imputed) (Table 2.6). The next lowest average of the random intercept predictions was for the partially observed/partially imputed intervals, -0.056; and the highest average of the random intercept predictions was for the completely observed intervals, 0.061 (Table 2.6). A similar trend for the average intercept predictions was observed for the MGMM with $\gamma = 1;$ mean = -0.168, -0.085, 0.047 respectively (Table 2.7). Positive average random intercept predictions indicate that on average, individual response profiles are higher than the population average. Similarly, negative average random intercept predictions indicate that on average, individual response profiles are lower than the population average.

Overall, the log-linear survey regression model parameter estimates were smaller for the MGMM imputation method versus the complete case analysis for $\gamma = 1$ and $\gamma = 1.083$. The standard errors associated with the model parameter estimates for the MGMM imputation method were appreciably lower, in most cases, than the corresponding standard errors produced by the complete case analysis (Table 2.4). All of the predictors for the complete case analysis were determined to be statistically significant at the alpha = 0.05 level. However, there was a marginal discrepancy between the complete case analysis and analyses based on data imputed using the MGMM method for Monday, Wednesday, Friday, and Saturday.

## 2.5 Discussion

Obtaining accurate measurements of physical activity and sedentary behavior from accelerometers is vital to understanding how they impact health outcomes of interest. While there are many advantages to using accelerometer data in epidemiologic studies, a traditional approach of summarizing data only for participants with a minimum number of adherent days each with a sufficient amount of wear time presents several challenges. First, this method assumes physical activity and sedentary behavior are MCAR during nonwear; however, sedentary behavior and physical activity are related to nonwear. Second, when summarizing at the day level, then averaging across days for each participant, variations in physical activity, sedentary behavior, and nonwear throughout a day are masked.

Generalized linear mixed models are appropriate for relating changes in the mean of a discrete response variable (e.g., counts per interval) to covariates.[23] These models typically apply a non-linear transformation to the mean of the discrete response variable which is related to a linear function of the covariates. The mixed model (i.e., fixed and random effects) framework was selected as it is ideal for modeling correlated, unbalanced, and hierarchically structured data. The introduction of random effects for each participant and day induces correlation among the multivariate count data, which yields accurate standard errors and thus valid inferences about the regression parameters can be obtained. Failure to account for the correlation and covariance among repeated measurements decreases the efficiency or precision with which the regression parameters can be estimated.[23] The nested structure of day within participant increases the precision with which regression parameters can be estimated.[23] The multilevel generalized mixed model for imputation can be implemented by most statistical software such as SAS, Stata, R, etc.

Even though we did not observe any meaningful differences between the weighted (for complex survey sampling designs) and unweighted analyses, we recommend performing both analyses as a general practice. This result is consistent with the findings from a simulation study comparing weighted and unweighted analyses of a multilevel linear mixed model for a continuous outcome (number of months without insurance).[55] The authors did not observe a difference in any inferential decisions between the weighted and unweighted analyses.[55] In addition to the GLLAMM package in R, multilevel mixed models for complex survey sampling data can also be constructed in Mplus and MLwiN.[55]

The accumulation of average counts/min by interval of the day varied between the complete case and MGMM methods (Table 2.4). In other words, if we ordered the intervals by decreasing average counts/min values, we would obtain a different sequence for the complete case and MGMM methods. For example, the highest average counts/min was observed for the 9:00 a.m.-12:00 p.m. interval for the complete case and MGMM ($\gamma = 1.083$) analyses but not for the MGMM ($\gamma = 1$) analysis. When this information was combined with the missing data percent for each interval displayed in Figure 2, there was no clear indication that intervals with smaller amounts of missing data had the highest average counts/min or that intervals with larger amounts of missing data had the lowest average counts/min. The first interval had the highest amount of missing data (22.3%) but was not the first or second interval with the lowest average counts/min (Figure 2). The accumulation of average counts/min by interval of the day was consistent between the complete case and GEE methods (Table 2.3). The disparity in the accumulation of average counts/min by interval of the day between the MGMM and GEE imputation methods can be partially explained by the distinct target of inference associated with generalized linear mixed models and marginal models (Section 4.1).[31] The target of inference for generalized linear mixed model is the individual and the population for marginal models.[23]

The multilevel generalized mixed model resulting in lower imputed values compared to the GEE could mechanistically be explained by negative values for the means of the random intercept predictions (Tables 2.6, 2.7). Random effects are a key and distinct feature of the multilevel generalized mixed model. On average, the subject-specific profiles were less than the population average profile. More precisely, the fixed intercept in the MGMM was less than the fixed intercept in the population average model by a factor of approximately $(\sigma_0^2 + \sigma_1^2)/2$.[31–33] Failing to include the estimate of $\gamma$ as a covariate in the imputation model may lead to an underestimation of counts/min and total counts per interval (Tables 2.3, 2.9). This was indicated by lower imputed values based on the models with $\gamma = 1$ versus $\gamma$ =1.083. Overall, we conclude that the rate of counts/min may not be the same during wear and nonwear periods.

This study had several limitations. First, a limitation of this approach could be obtaining imputed values for large study populations. The computational intensity of obtaining subject-specific intercepts, particularly with nested effects, increases with increased sample sizes. Sufficient computing resources are necessary for this procedure. Second, the analysis sample was comprised of Hispanic/Latino adults, which could limit generalizability. Other study populations may have a larger amount of missing data or exhibit a different missing data pattern. Third, there was some indication of the potential benefits of the MGMM imputation method, however, it will be important to assess its performance using a simulation study.

Imputation of missing accelerometer data is often considered an improvement upon the ad hoc approach of excluding participants with insufficient data. Even though several investigators have introduced promising imputation procedures, a consensus on the optimal way to summarize accelerometer data that contains missing values has yet to be established by the research community. The variety of the proposed imputation techniques makes it challenging to compare the performance of the techniques across studies. Measures of imputation bias and precision are useful in determining the performance of an imputation procedure. Simulations with varying amounts of missing data combined with different missing data mechanisms should be conducted in the future in order to assess the performance of the multilevel generalized mixed model imputation method.

## 2.6 Tables and Figures

Table 2.1. Weighted descriptive characteristics for the full study population, n=16,415, and the analysis subset, n=1,000, HCHS/SOL 2008-2011

| Characteristic | Total n | Weighted Percent | Total n | Weighted Percent |
|---|---|---|---|---|
| Overall | 16,415 | 100.0 | 1,000 | 7.1 |
| Gender: | | | | |
| Female | 9,835 | 52.1 | 580 | 50.8 |
| Male | 6,580 | 47.9 | 420 | 49.3 |
| Age (yr): | | | | |
| 18-24 | 1,665 | 16.8 | 82 | 13.6 |
| 25-34 | 2,082 | 21.8 | 105 | 20.1 |
| 35-44 | 2,954 | 21.2 | 167 | 20.5 |
| 45-54 | 4,922 | 18.9 | 323 | 20.4 |
| 55-64 | 3,460 | 12.8 | 236 | 16.3 |
| 65+ | 1,332 | 8.5 | 87 | 9.1 |
| Body mass index ($kg/m^2$): | | | | |
| Underweight | 130 | 1.2 | 10 | 1.4 |
| Normal | 3,191 | 22.1 | 199 | 20.1 |
| Overweight | 6,116 | 37.2 | 370 | 40.5 |
| Obese | 6,907 | 39.6 | 420 | 38.1 |

Table 2.2. Imputation model parameter estimates and standard errors for total counts per interval based on complete cases using a multilevel generalized linear mixed model (MGMM) with offset ($\gamma = 1$) and with log wear time as covariate ($\gamma = 1.083$), n=1000 participants (35,982 intervals/observations)

| | MGMM $\gamma = 1$ | MGMM $\gamma = 1.083$ |
|---|---|---|
| Variable | Estimate (SE) | Estimate (SE) |
| $\log(N_{ijt})$ | 1.000 (--) | 1.083 (0.034) |
| Intercept | 5.792 (0.172) | 4.788 (0.279) |
| Age (yr) | -0.015 (0.002) | -0.007 (0.003) |
| Male | 0.268 (0.056) | 0.267 (0.059) |
| BMI (kg/m$^2$) | -0.016 (0.004) | -0.010 (0.003) |
| Day of Week | | |
| Monday | 0.219 (0.032) | 0.215 (0.031) |
| Tuesday | 0.202 (0.033) | 0.198 (0.032) |
| Wednesday | 0.178 (0.031) | 0.172 (0.031) |
| Thursday | 0.244 (0.032) | 0.239 (0.031) |
| Friday | 0.216 (0.031) | 0.211 (0.031) |
| Saturday | 0.115 (0.031) | 0.112 (0.031) |
| Interval | | |
| 9:00am-12:00pm | 0.122 (0.041) | 0.110 (0.042) |
| 12:00pm-3:00pm | 0.101 (0.044) | 0.086 (0.046) |
| 3:00pm-6:00pm | 0.006 (0.046) | -0.009 (0.048) |
| 6:00pm-9:00pm | -0.191 (0.049) | -0.203 (0.050) |
| 9:00pm-12:00am | -0.515 (0.062) | -0.516 (0.062) |
| Var($b_{0i}$) = $\sigma_0^2$ | 0.613 (0.285) | 0.610 (0.309) |
| Var($b_{1ij}$) = $\sigma_1^2$ | 0.345 (0.016) | 0.342 (0.016) |

Survey design strata were included in the imputation models to account for the complex survey design. Z – Statistic and 95% CI for test of $\gamma \neq 1$; 2.441 (1.017, 1.150). MGMM = multilevel generalized mixed model.

Table 2.3. Average counts/min per interval based on MGMM and GEE imputation models and 5 imputations.

| Characteristic | CCA 12% missing | MGMM $\gamma = 1.083$ | MGMM $\gamma = 1$ | GEE $\gamma = 1.202$ | GEE $\gamma = 1$ |
|---|---|---|---|---|---|
| Overall | 176.43 | 171.18 | 158.20 | 179.95 | 178.63 |
| Sex | | | | | |
| Male | 208.08 | 202.650 | 182.529 | 212.466 | 211.795 |
| Female | 153.50 | 148.423 | 133.329 | 156.431 | 154.642 |
| Age (yr) | | | | | |
| 18-24 | 230.90 | 228.266 | 201.003 | 245.826 | 244.315 |
| 25-34 | 205.01 | 198.529 | 178.859 | 213.151 | 210.022 |
| 35-44 | 204.51 | 195.201 | 175.361 | 202.608 | 200.777 |
| 45-54 | 174.38 | 165.978 | 150.091 | 174.914 | 173.432 |
| 55-64 | 152.53 | 149.401 | 135.093 | 154.135 | 154.107 |
| 65+ | 115.56 | 116.652 | 104.227 | 122.982 | 122.107 |
| Interval | | | | | |
| 6:00am-9:00am | 189.16 | 177.745 | 144.911 | 200.212 | 195.908 |
| 9:00am-12:00pm | 206.06 | 205.721 | 187.704 | 212.795 | 212.518 |
| 12:00pm-3:00pm | 203.01 | 202.994 | 191.489 | 206.188 | 206.043 |
| 3:00pm-6:00pm | 186.32 | 184.903 | 174.903 | 188.630 | 188.284 |
| 6:00pm-9:00pm | 154.16 | 150.797 | 139.163 | 155.823 | 155.409 |
| 9:00pm-12:00am | 114.25 | 104.936 | 85.703 | 116.046 | 113.616 |
| Day of Week | | | | | |
| Sunday | 148.00 | 144.286 | 123.607 | 152.851 | 150.695 |
| Monday | 184.30 | 177.580 | 160.197 | 188.985 | 187.768 |
| Tuesday | 181.55 | 175.395 | 157.829 | 184.172 | 183.202 |
| Wednesday | 177.90 | 176.260 | 159.591 | 180.704 | 179.952 |
| Thursday | 192.86 | 188.483 | 172.060 | 195.191 | 194.309 |
| Friday | 184.16 | 176.765 | 163.088 | 188.542 | 187.219 |
| Saturday | 166.01 | 162.199 | 144.637 | 171.651 | 169.800 |
| Weekday | 184.20 | 178.895 | 162.542 | 187.582 | 186.549 |
| Weekend day | 156.30 | 152.268 | 132.978 | 161.228 | 159.208 |

N = 1,000 participants. Twelve percent of all 35,982 intervals were entirely missing. CCA = complete case analysis (i.e., participants with ≥10 h of wear time for ≥3 d), MGMM = multilevel generalized mixed model, and GEE = generalized estimating equations.

Table 2.4. Log-linear (Poisson) survey regression model parameter estimates (SE) for average counts/min per interval using CCA and MGMM based on 5 imputations

| Variable | CCA | MGMM $\gamma = 1.083$ | MGMM $\gamma = 1$ |
|---|---|---|---|
| Intercept | 9.047 (0.155) | 9.182 (0.146) | 9.312 (0.142) |
| Age (y) | -0.009 (0.002) | -0.008 (0.002) | -0.009 (0.002) |
| Male | 0.255 (0.065) | 0.255 (0.059) | 0.257 (0.058) |
| BMI (kg/m$^2$) | -0.012 (0.004) | -0.013 (0.004) | -0.014 (0.004) |
| Day of Week | | | |
| Monday | 0.226 (0.048) | 0.141 (0.073) | 0.145 (0.071) |
| Tuesday | 0.278 (0.038) | 0.216 (0.067) | 0.221 (0.064) |
| Wednesday | 0.200 (0.042) | 0.133 (0.071) | 0.138 (0.068) |
| Thursday | 0.222 (0.045) | 0.166 (0.073) | 0.168 (0.071) |
| Friday | 0.200 (0.038) | 0.102 (0.064) | 0.108 (0.061) |
| Saturday | 0.163 (0.044) | 0.091 (0.068) | 0.098 (0.064) |
| Interval | | | |
| 9:00am-12:00pm | 0.246 (0.049) | 0.115 (0.029) | 0.094 (0.028) |
| 12:00pm-3:00pm | 0.285 (0.063) | 0.118 (0.039) | 0.086 (0.038) |
| 3:00pm-6:00pm | 0.241 (0.068) | 0.070 (0.045) | 0.035 (0.044) |
| 6:00pm-9:00pm | 0.063 (0.072) | -0.113 (0.046) | -0.145 (0.045) |
| 9:00pm-12:00am | -0.344 (0.075) | -0.458 (0.046) | -0.473 (0.045) |

N = 1,000 participants. Reference levels are Female, Sunday and 6:00 a.m.-9:00 a.m. CCA = complete case analysis (i.e., participants with ≥10 h of wear time for ≥3 d), MGMM = multilevel generalized mixed model.

Table 2.5. Imputation model parameter estimates and standard errors (SE) for total counts per interval based on complete cases using GEE and offset ($\gamma$)

| | GEE $\gamma = 1$ | GEE $\gamma = 1.202$ |
|---|---|---|
| Variable | Estimate (SE) | Estimate (SE) |
| $\log(N_{ijt})$ | 1.000 (--) | 1.202 (0.034) |
| Intercept | 6.408 (0.095) | 5.402 (0.194) |
| Age (yr) | -0.010 (0.001) | -0.010 (0.001) |
| Male | 0.268 (0.016) | 0.267 (0.016) |
| BMI (kg/m²) | -0.012 (0.001) | -0.012 (0.001) |
| Day of Week | | |
| Monday | 0.219 (0.029) | 0.215 (0.029) |
| Tuesday | 0.202 (0.030) | 0.198 (0.030) |
| Wednesday | 0.177 (0.030) | 0.172 (0.030) |
| Thursday | 0.244 (0.029) | 0.239 (0.029) |
| Friday | 0.216 (0.030) | 0.211 (0.030) |
| Saturday | 0.115 (0.031) | 0.112 (0.031) |
| Interval | | |
| 9:00am-12:00pm | 0.081 (0.027) | 0.052 (0.028) |
| 12:00pm-3:00pm | 0.050 (0.027) | 0.014 (0.028) |
| 3:00pm-6:00pm | -0.040 (0.027) | -0.076 (0.028) |
| 6:00pm-9:00pm | -0.232 (0.029) | -0.262 (0.029) |
| 9:00pm-12:00am | -0.545 (0.035) | -0.547 (0.035) |

N = 1,000 participants. Survey design strata were included in the imputation models to account for the complex survey sampling design. GEE = generalized estimating equations. Z – Statistic and 95% CI for test of $\gamma = 1$; 5.941 (1.136, 1.268).

Table 2.6. Descriptive characteristics for the random intercept predictions in the MGMM for average counts/min per interval by imputation method, $\gamma = 1.083$.

| | n | Variable | Mean | SD | Minimum | Maximum |
|---|---|---|---|---|---|---|
| Completely observed intervals | 21,208 | $\overline{\widehat{b_{0i}}}$ | 0.031 | 0.554 | -2.281 | 2.163 |
| | | $\overline{\widehat{b_{1ij}}}$ | 0.029 | 0.448 | -3.270 | 2.088 |
| | | $\overline{(\widehat{b_{0i}} + \widehat{b_{1ij}})}$ | 0.061 | 0.733 | -3.819 | 3.271 |
| Entirely missing intervals | 4,230 | $\overline{\widehat{b_{0i}}}$ | 0.007 | 0.519 | -2.067 | 1.595 |
| | | $\overline{\widehat{b_{1ij}}}$ | -0.076 | 0.754 | -3.478 | 2.767 |
| | | $\overline{(\widehat{b_{0i}} + \widehat{b_{1ij}})}$ | -0.069 | 0.976 | -3.636 | 3.318 |
| Partially observed/partially imputed interval | 10,544 | $\overline{\widehat{b_{0i}}}$ | -0.030 | 0.576 | -2.281 | 2.163 |
| | | $\overline{\widehat{b_{1ij}}}$ | -0.026 | 0.585 | -3.589 | 2.767 |
| | | $\overline{(\widehat{b_{0i}} + \widehat{b_{1ij}})}$ | -0.056 | 0.872 | -4.443 | 3.318 |

N = 1,000 participants. $\overline{\widehat{b_{0i}}}$ is the average of the subject-specific intercepts. $\overline{\widehat{b_{1ij}}}$ is the average of the random intercepts for each day *j* nested within subject *i*. MGMM = multilevel generalized mixed model.

Table 2.7. Descriptive characteristics for the random intercept predictions in the MGMM for average counts/min per interval by imputation method, $\gamma = 1$.

| | n | Variable | Mean | SD | Minimum | Maximum |
|---|---|---|---|---|---|---|
| Completely observed intervals | 21,208 | $\overline{\widehat{b_{0i}}}$ | 0.011 | 0.559 | -2.339 | 2.264 |
| | | $\overline{\widehat{b_{1ij}}}$ | 0.036 | 0.450 | -3.271 | 2.097 |
| | | $\overline{(\widehat{b_{0i}} + \widehat{b_{1ij}})}$ | 0.047 | 0.735 | -3.849 | 3.228 |
| Entirely missing intervals | 4,230 | $\overline{\widehat{b_{0i}}}$ | -0.046 | 0.535 | -2.067 | 1.592 |
| | | $\overline{\widehat{b_{1ij}}}$ | -0.122 | 0.770 | -3.648 | 2.731 |
| | | $\overline{(\widehat{b_{0i}} + \widehat{b_{1ij}})}$ | -0.168 | 1.010 | -4.025 | 3.228 |
| Partially observed/partially imputed interval | 10,544 | $\overline{\widehat{b_{0i}}}$ | -0.055 | 0.576 | -2.339 | 2.264 |
| | | $\overline{\widehat{b_{1ij}}}$ | -0.029 | 0.595 | -3.648 | 2.731 |
| | | $\overline{(\widehat{b_{0i}} + \widehat{b_{1ij}})}$ | -0.085 | 0.880 | -4.526 | 3.228 |

N = 1,000 participants. $\overline{\widehat{b_{0i}}}$ is the average of the subject-specific intercepts. $\overline{\widehat{b_{1ij}}}$ is the average of the random intercepts for each day *j* nested within subject *i*. MGMM = multilevel generalized mixed model.

Table 2.8. Survey regression model parameter estimates (SE) for average counts/min per interval using CCA and GEE based on 5 imputations

| Variable | CCA | GEE $\gamma = 1.202$ | GEE $\gamma = 1$ |
|---|---|---|---|
| Intercept | 9.047 (0.155) | 9.314 (0.117) | 9.311 (0.117) |
| Age (yr) | -0.009 (0.002) | -0.008 (0.002) | -0.008 (0.001) |
| Male | 0.255 (0.065) | 0.263 (0.056) | 0.262 (0.055) |
| BMI (kg/m$^2$) | -0.012 (0.004) | -0.014 (0.004) | -0.014 (0.003) |
| Day of Week | | | |
| Monday | 0.226 (0.048) | 0.214 (0.036) | 0.216 (0.036) |
| Tuesday | 0.278 (0.038) | 0.244 (0.029) | 0.247 (0.029) |
| Wednesday | 0.200 (0.042) | 0.166 (0.032) | 0.169 (0.031) |
| Thursday | 0.222 (0.045) | 0.201 (0.036) | 0.204 (0.036) |
| Friday | 0.200 (0.038) | 0.172 (0.029) | 0.176 (0.029) |
| Saturday | 0.163 (0.044) | 0.123 (0.031) | 0.125 (0.031) |
| Interval | | | |
| 9:00am-12:00pm | 0.246 (0.049) | 0.036 (0.027) | 0.044 (0.027) |
| 12:00pm-3:00pm | 0.285 (0.063) | 0.025 (0.037) | 0.030 (0.037) |
| 3:00pm-6:00pm | 0.241 (0.068) | -0.027 (0.043) | -0.023 (0.043) |
| 6:00pm-9:00pm | 0.063 (0.072) | -0.200 (0.045) | -0.193 (0.045) |
| 9:00pm-12:00am | -0.344 (0.075) | -0.497 (0.044) | -0.494 (0.044) |

N = 1,000 participants. Reference levels are Female, Sunday and 6:00 a.m.-9:00 a.m. CCA = complete case analysis (i.e., participants with ≥10 h of wear time for ≥3 d), GEE = generalized estimating equations.

Table 2.9. Average (SD) total counts per interval using the MGMM and GEE imputation models and 5 imputations.

| Characteristic | MGMM $\gamma = 1.083$ | MGMM $\gamma = 1$ | GEE $\gamma = 1.202$ | GEE $\gamma = 1$ |
|---|---|---|---|---|
| Overall | 30,813 (43,893) | 27,716 (42,783) | 32,391 (41,760) | 32,153 (41,772) |
| Sex | | | | |
| Male | 36,477 (52,089) | 32,855 (50,934) | 38,244 (49,632) | 38,123 (49,673) |
| Female | 26,716 (36,303) | 23,999 (35,286) | 28,158 (34,365) | 27,836 (34,323) |
| Age (yr) | | | | |
| 18-24 | 41,088 (60,496) | 36,181 (57,413) | 44,249 (55,668) | 43,977 (55,686) |
| 25-34 | 35,735 (47,024) | 32,195 (46,389) | 38,367 (45,099) | 37,803 (45,097) |
| 35-44 | 35,136 (49,974) | 31,565 (48,074) | 36,469 (46,331) | 36,139 (46,377) |
| 45-54 | 29,878 (38,537) | 27,016 (37,970) | 31,484 (36,906) | 31,218 (36,951) |
| 55-64 | 26,892 (40,086) | 24,317 (39,647) | 27,744 (38,833) | 27,739 (38,865) |
| 65+ | 20,997 (32,500) | 18,761 (32,400) | 22,136 (32,061) | 21,979 (32,012) |
| Interval | | | | |
| 6:00am-9:00am | 31,994 (43,581) | 26,084 (41,884) | 36,038 (40,367) | 35,263 (40,378) |
| 9:00am-12:00pm | 37,029 (46,729) | 33,787 (45,185) | 38,303 (43,907) | 38,253 (43,908) |
| 12:00pm-3:00pm | 36,539 (46,414) | 34,468 (44,946) | 37,113 (44,073) | 37,088 (44,082) |
| 3:00pm-6:00pm | 33,283 (43,566) | 31,482 (42,583) | 33,953 (41,801) | 33,891 (41,810) |
| 6:00pm-9:00pm | 27,143 (41,162) | 25,049 (40,431) | 28,048 (39,702) | 27,974 (39,694) |
| 9:00pm-12:00am | 18,888 (38,699) | 15,427 (38,239) | 20,888 (37,650) | 20,451 (37,643) |
| Day of Week | | | | |
| Sunday | 25,972 (37,706) | 22,249 (35,086) | 27,513 (33,817) | 27,125 (33,801) |
| Monday | 31,964 (46,712) | 28,836 (46,036) | 34,017 (45,041) | 33,798 (45,061) |
| Tuesday | 31,571 (41,011) | 28,409 (40,073) | 33,151 (39,092) | 32,976 (39,134) |
| Wednesday | 31,727 (48,292) | 28,726 (46,029) | 32,526 (44,809) | 32,391 (44,821) |
| Thursday | 33,927 (49,519) | 30,971 (48,752) | 35,134 (47,750) | 34,976 (47,765) |
| Friday | 31,818 (43,305) | 29,356 (43,263) | 33,938 (42,554) | 33,699 (42,558) |
| Saturday | 29,196 (39,063) | 26,035 (38,410) | 30,897 (37,489) | 30,564 (37,459) |
| Weekday | 32,201 (45,891) | 29,258 (44,954) | 33,765 (43,973) | 33,579 (43,990) |
| Weekend day | 27,408 (38,348) | 23,936 (36,651) | 29,021 (35,538) | 28,657 (35,517) |

N = 1,000 participants. MGMM = multilevel generalized mixed model. GEE = generalized estimating equations.

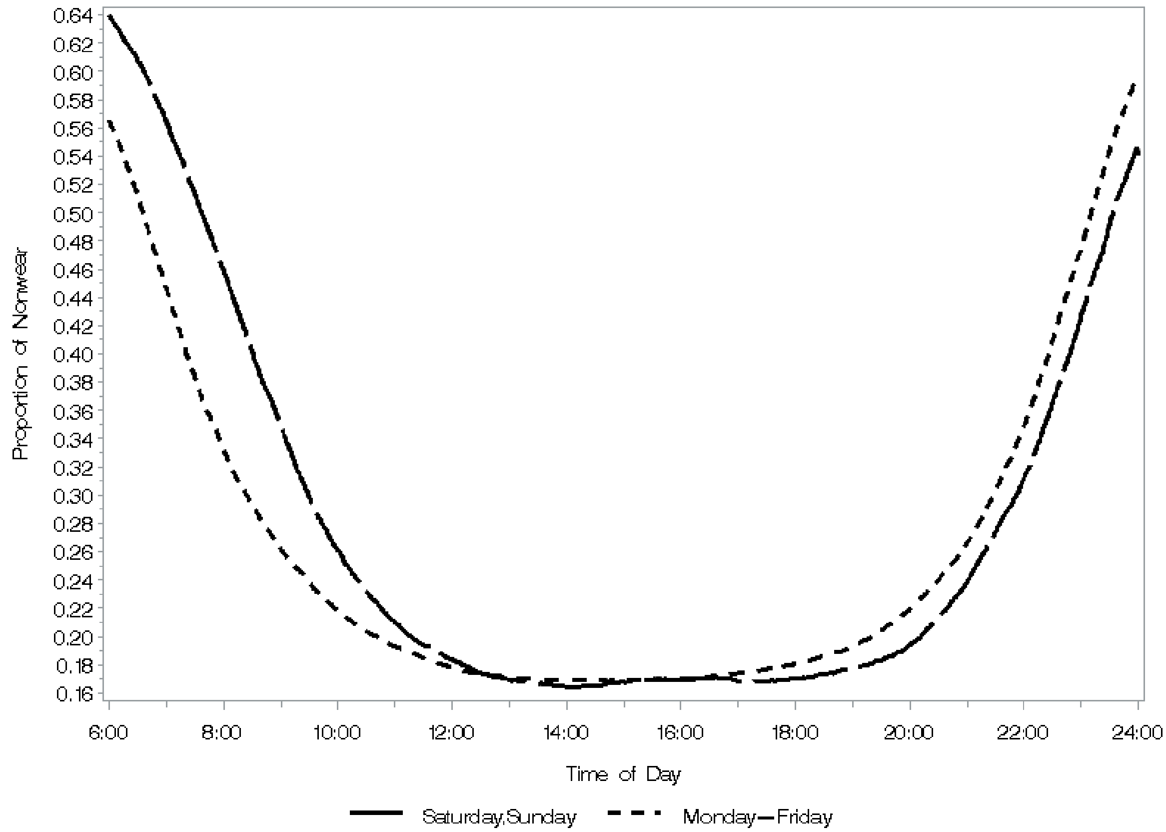


Figure 2.1. Proportion of participants not wearing the accelerometer (i.e., registering zero count activity) by time of day and day of week.

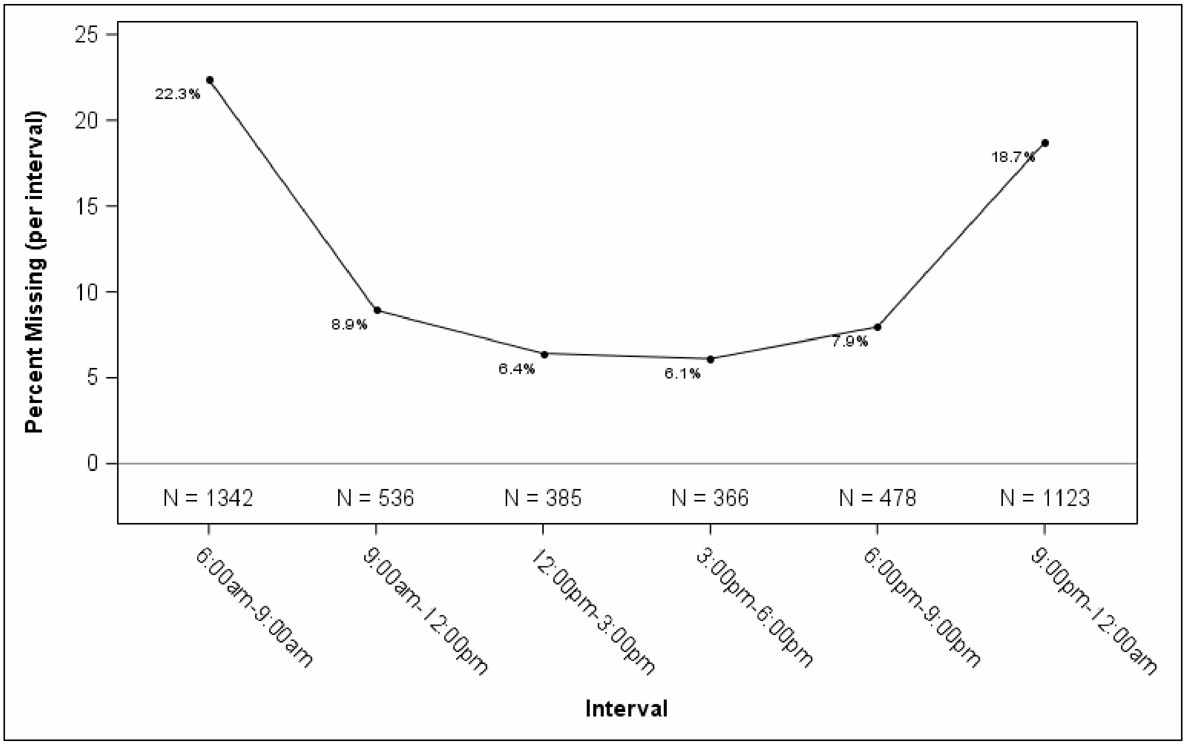

Figure 2.2. Frequency (n) and percent of missing data per interval. There were 5,997 observations in each interval.

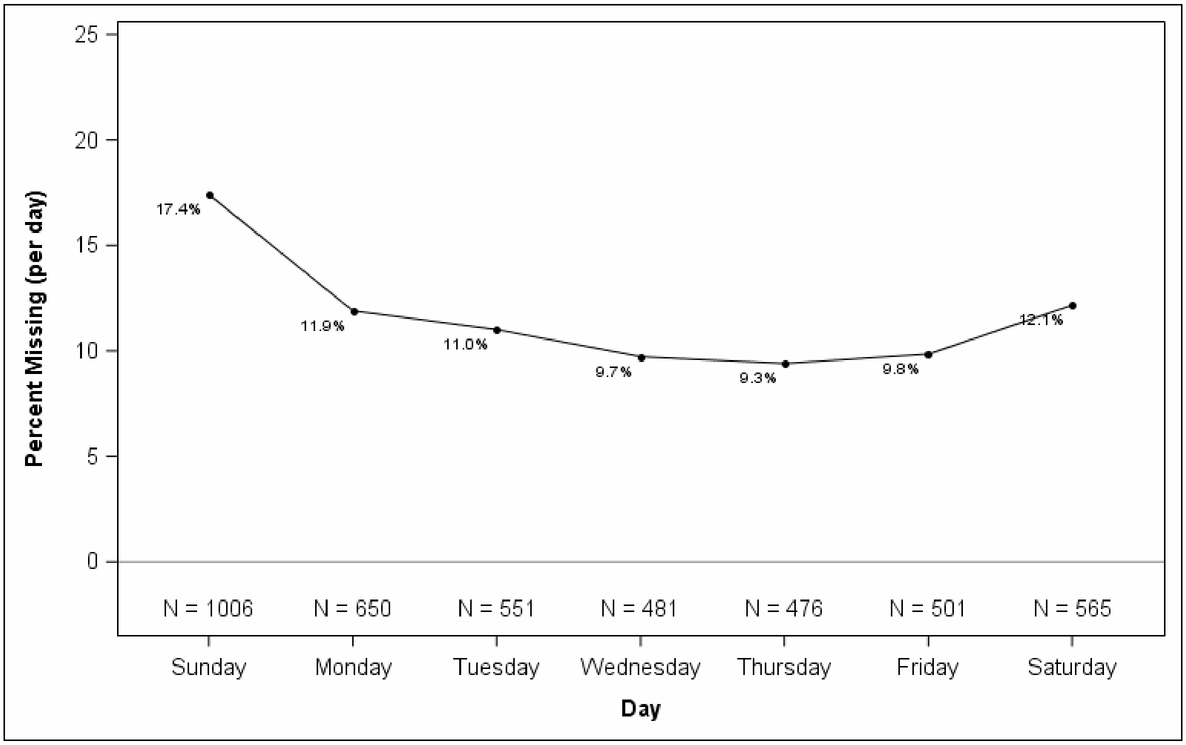

Figure 2.3. Frequency (n) and percent of missing intervals per day (6:00 a.m.-12:00 a.m). The number of intervals on each day ranged from 4,644 on Saturday to 5,778 on Sunday.

# CHAPTER 3: A SIMULATION STUDY TO ASSESS THE PERFORMANCE OF THE MULTILEVEL GENERALIZED MIXED MODEL METHOD USED TO IMPUTE ACCELEROMETER MISSING DATA IN THE HCHS/SOL STUDY

## 3.1 Introduction

A physically active lifestyle among children and adults promotes short- and long-term physical and mental well-being.[35] They are less likely to become obese and develop related chronic illnesses such as type 2 diabetes, cardiovascular disease, depression, and some cancers.[1,3,35,36] Accelerometry is commonly used to measure physical activity in large epidemiologic studies. Strengths of accelerometry include the accuracy with which the device is able to capture physical activity and the ability to record large amounts of data. One limitation of hip-worn devices is that they do not capture upper-body movement well.[6] They primarily measure locomotor activity.[6] Lastly, accurately and precisely summarizing accelerometer data in the presence of missing data creates a variety of analytical challenges. The challenges of analyzing counts data is only applicable to certain accelerometers that apply a proprietary conversion algorithm to the raw data.

Devices, such as the Phillips Respironics Actical accelerometer, worn on the hip, can be used to estimate physical activity and sedentary behavior. These devices measure acceleration in units of gravity ($g \approx 9.81$ m/s$^2$) at a specific frequency (e.g., 30 times/s or 30 Hz).[10] The company aggregates accelerations to activity counts per user-specified unit of time, known as an epoch.[10]

A challenge in free-living accelerometer studies where the data is only available through a proprietary algorithm (e.g., Actical) is distinguishing sedentary behavior from nonwear since theoretically, both can register zero counts/epoch (e.g., 0 counts/60 s epoch).[7] One classification of sedentary behavior is defined as 0 to 100 counts/min.[12] This classification is sensitive to the nonwear definition because a low or conservative threshold for identifying nonwear time is more likely to classify sedentary periods as nonwear.[38,39] Nonwear time, resulting in a period of consecutive zero counts, later becomes missing data which can bias assessments of physical activity.[21] Estimates of physical activity are biased downward as counts are not recorded during nonwear.[9] To circumvent this potential bias, researchers sometimes analyze data only for participants with a minimum number of adherent days, defined as having a sufficient amount of wear time in a given day (ad hoc approach).[9,12] Non-adherent days are labelled as "missing". The concern with this approach is that it assumes physical activity and sedentary behavior are missing completely at random (MCAR) during nonwear. For example, the MCAR assumption would be violated if participants are more (or less) likely to wear the device during nonwear.[9] A number of researchers have had success with reducing bias in estimates of physical activity and sedentary behavior by implementing missing data approaches that incorporate data from adherent and non-adherent days instead of excluding that information as the ad hoc approach does.[9,21]

Simulations studies are commonly employed as a means to assess the performance of an imputation procedure using some measure of bias. Smaller values of bias are indicative of greater accuracy. Catellier et al. implemented a single value imputation method, the expectation-maximization algorithm, and multiple imputation (MI) to impute missing accelerometer data for the entire day and interval of the day.[9] The simulation for this study showed that the proposed method successfully reduced bias, based on the mean signed difference, and that the performance of each imputation method depended on the proportion of missing data, the correlation of activity across days of the week, and the missing data mechanism.[9] Using data from adherent and non-adherent days, Xu proposed a mixed model technique for imputation, which accounts for variation between and within participants.[21] This imputation method was shown to successfully reduce bias comparing true and imputed values for model parameter estimates.[21] Assessments of accelerometer missing data approaches are essential for determining the reliability of estimates of physical activity and sedentary behavior produced by these approaches.

Reliable estimates of physical activity and sedentary behavior, given missing data, are key to understanding how the total volume and patterns of accumulation of physical activity and sedentary behavior affect health outcomes.[40,44] Average counts/min is an indicator of average total volume of physical activity and was the main outcome of interest for this study.

In Chapter 2 we presented a statistical method to impute missing accelerometer data that incorporates the use of all available data, variability from within and between participants, and is well suited to account for multivariate count data under a complex survey design; the multilevel generalized mixed model (MGMM). Our results indicated that failure to include accelerometer wear time as a predictor, instead of as an exposure variable in the multilevel generalized mixed model for imputation, may lead to an underestimation of accelerometer average counts per minute. Moreover, the rate of average counts per minute may not be the same during wear and nonwear periods which provides evidence against the assumption for missing completely at random in the ad hoc approach.

The purpose of this simulation study is to assess the performance of the proposed method using a measure of percent relative bias. Smaller values of this metric are indicative of greater accuracy. We hypothesize that percent relative bias will be smaller for MGMM imputation evaluation models that are congruent with MGMM data generation models. For example, when counts per interval data are generated using the multilevel Poisson mixed model with the coefficient of the offset equal to one, the multilevel Poisson mixed model with the coefficient of the offset equal to one will yield the smallest estimates of percent relative bias. In other words, evaluation models that are accurately specified will produce lower estimates of percent relative bias.

## 3.2 Physical Activity Demonstration Example

Accelerometer data from The Hispanic Community Health Study/Study of Latinos, 2008 – 2011, was used to demonstrate the methodology and to generate data for the simulation.

### 3.2.1 Study Population

The Hispanic Community Health Study/Study of Latinos (HCHS/SOL) is a multicenter community-based cohort study designed to examine the risk factors of chronic disease among Hispanics/Latinos, aged 18 to 74, in the United States.[45] The HCHS/SOL enrolled 16,415 self-identified Hispanics/Latinos between March 2008 and June 2011 in four communities (the Bronx, New York; Chicago, Illinois; Miami, Florida; San Diego, California) in the United States. Recruitment was implemented through a two-stage area household probability sampling design.[45,46] Further details on the study design have been previously published.[45,46]

### 3.2.2 Data Measures

Accelerometer data was obtained from participants at the four HCHS/SOL field centers. During the baseline clinic visit, participants were asked to wear the Actical™ accelerometer (model 198-0200-03, Minimiter Respironics®, Bend, OR) for seven days during waking hours. They were instructed to undertake usual activities while wearing the monitor on the hip and to remove it only for swimming, bathing, and sleeping. In addition to receiving instructions on proper wear during the clinic visit, participants were given written instructions and a phone number to call if questions arose during the 7-day monitoring period. At the end of the monitoring period, participants returned the Actical to the field center in person or via mail.

To standardize across sites, data recorded beginning at 5:00 a.m. on the day after the clinic visit, ending at a maximum of six days, were included. The length of day one was 19 hours (5:00 a.m.-11:59 p.m.) and the length of days two through six was 24 hours (12:00 a.m.-11:59 p.m.).[47]

The Actical accelerometer measures movement in all directions and was programmed to record counts and steps in 60-s epochs. Average counts/min was the primary measure for this study. Nonwear time was defined as at least 90 minutes of consecutive zero counts, with the allowance for intervals of up to two minutes of nonzero counts if the 30 minutes preceding and following these intervals were consecutive zero counts.[48]

### 3.2.3 Analytic Sample

A total of 16,415 participants were recruited for the original study. There were 12,750 participants classified as adherent (i.e., ≥10 h of wear time for ≥3 d). Of the adherent participants, 608 were excluded based on reporting having worked the night shift (after midnight) occasionally or regularly .[47] Moreover, only data recorded between 6:00 a.m. and 12:00 midnight each day (18 h) were included. Our interest was in estimating physical activity when participants have been instructed to wear the device (i.e., during waking hours) thus the start of our monitoring day at 6:00 a.m. Similar to Lee (2013), data from non-adherent (or invalid) days (i.e., <10 h/d) for adherent participants were included instead of excluding them as the ad hoc approach does. A group of 1,000 participants with a moderate amount of missing data (including intervals with both wear- and nonwear time, as well as intervals that are completely missing, i.e., all nonwear time) were randomly selected from the group of 12,750 adherent participants. Approximately 34% of the intervals had some nonwear time.

### 3.2.4 Physical Activity Outcome

Each 18-h monitoring day was divided into six intervals of equal length: (1) 6:00 a.m.-9:00 a.m., (2) 9:00 a.m.-12:00 p.m., (3) 12:00 p.m.-3:00 p.m., (4) 3:00 p.m.-6:00 p.m., (5) 6:00 p.m.-9:00 p.m., and (6) 9:00 p.m.-12:00 midnight. The maximum possible wear time for each interval was three hours (180 min). We used six intervals of equal length for illustration purposes. Nevertheless, the length of the interval was arbitrary and could have been different. The physical activity outcome of interest was average counts/min in an interval, an indicator of average total volume of physical activity in an interval. This was calculated by taking the sum of all counts during wear time in the interval and dividing by the total wear time (in minutes) in that interval. Log-linear (Poisson) regression enables conversion of the dependent variable from a rate into a count. The log of the rate, average counts/min, is modeled as a function of a set of covariates in the multilevel generalized mixed model such that log(counts/min) = log(counts) – log(min) on the left-hand-side. The conversion of the dependent variable from a rate into a count is obtained by adding the log(min) term to the right-hand-side as an offset (or covariate) leaving log(counts) on the left-hand-side.

## 3.3 Overview of Statistical Analyses

The multilevel generalized mixed model (MGMM) method takes advantage of the richness in information provided by all available data among those with at least a certain number of adherent days. Under the assumption of missing at random (MAR), parameter estimates from the MGMM of counts per interval as a function of wear time, age (years), body mass index (BMI) ($kg/m^2$), sex, day (Sunday, Monday,…, Saturday), and interval of the day were used to obtain unknown counts per interval during nonwear. To assess the performance of this method a simulation study was conducted as follows:

### 3.3.1 Methods

### 3.3.2 Steps for Data Creation

**1.** One thousand participants with a moderate amount of missing data were identified where "moderate" was defined as at least 30% of intervals have some nonwear time. This included intervals that were entirely missing (i.e., all nonwear time), partially missing, or fully observed. Descriptive information for this group of participants as well as the frequency and type of intervals (i.e., entirely missing, partially missing, fully observed) are reported.

**2.** A Poisson random effects model was fit using SAS software 9.4 (SAS Institute, Cary, NC) to the observed data where $Y_{ijk}$ denotes the observed counts per interval during wear time ($w_{ijk}$), $\sigma_0^2$ is the variance of the subject random intercept $b_{0i}$, and $\sigma_1^2$ is the variance for the day random intercept $b_{1ij}$. Random effects were estimated using empirical Bayes estimation. Wear time was specified as a predictor variable for which the associated model coefficient $\gamma$ was estimated. Age (years), BMI (kg/m$^2$), and wear time minutes were continuous; sex, day of the week (Sunday, Monday,…,Saturday), and interval of the day were categorical. Intervals were nested within day and day within subject, hence the multilevel structure.

All model parameter estimates were retained except for $\gamma$. Instead of using the estimated value for $\gamma$, we investigate the cases where $\gamma = 1.3$ and $\gamma = 1.0$.

$$log\left(E\left(\frac{Y_{ijk}}{w_{ijk}^{\gamma}}\middle|\boldsymbol{b_i}\right)\right) = log\left(\frac{\mu_{ijk}^{w}}{w_{ijk}^{\gamma}}\right);$$

$$\begin{aligned} log(\mu_{ijk}^{w}) &= \gamma log(w_{ijk}) + \beta_0 + \beta_1 Age_i + \beta_2 Sex_i + \beta_3 BMI_i + \beta_4 I(Day_{ij} = Monday) \\ &+ \cdots + \beta_9 I(Day_{ij} = Saturday) + \beta_{10} I(Interval_{ijk} = 9am - 12pm) \\ &+ \cdots + \beta_{14} I(Interval_{ijk} = 9pm - 12am) + b_{0i} + b_{1ij} \end{aligned}$$

**3.** For the case of $\gamma = 1.3$, $r = 1, \ldots, 200$ replicate data sets were created with 1,000 unique participants and 35,988 total intervals. Wear time ($w_{ijk}$) and nonwear time ($w_{ijk}^{nw} = 180 - w_{ijk}$) as observed in original data were the same across all 200 replicate data sets. In each replicate data set, (i) we randomly drew $b_{0i}^{(r)} \sim N(0, \sigma_0^2)$ for each participant, $i = 1, \ldots, 1{,}000$ and (ii) randomly drew $b_{1ij}^{(r)} \sim N(0, \sigma_1^2)$. Next, for each interval $k$ on day $j$ for participant $i$ (iii) we calculated the

(a) Conditional wear means $\mu_{ijk}^{w(r)}$ given $b_{0i}^{(r)}$, $b_{1ij}^{(r)}$, $w_{ijk}$, $\boldsymbol{\beta}$, and $\gamma = 1.3$; and

(b) Conditional nonwear means $\mu_{ijk}^{nw(r)}$ given $b_{0i}^{(r)}$, $b_{1ij}^{(r)}$, $w_{ijk}^{nw}$, $\boldsymbol{\beta}$, and $\gamma = 1.3$; and randomly drew

(c) the wear count $Y_{ijk}^{w} \sim Poisson(\mu_{ijk}^{w})$ and (d) the nonwear count $Y_{ijk}^{nw} \sim Poisson(\mu_{ijk}^{nw})$.

$$log(\hat{\mu}_{ijk}^{w}) = 1.3log(w_{ijk}) + \hat{\beta}_0 + \hat{\beta}_1 Age_i + \hat{\beta}_2 Sex_i + \hat{\beta}_3 BMI_i + \hat{\beta}_4 I(Day_{ij} = Monday)$$
$$+ \cdots + \hat{\beta}_9 I(Day_{ij} = Saturday) + \hat{\beta}_{10} I(Interval_{ijk} = 9am - 12pm)$$
$$+ \cdots + \hat{\beta}_{14} I(Interval_{ijk} = 9pm - 12am + \hat{b}_{0i} + \hat{b}_{1ij})$$

$$log(\hat{\mu}_{ijk}^{nw}) = 1.3log(w_{ijk}^{nw}) + \hat{\beta}_0 + \hat{\beta}_1 Age_i + \hat{\beta}_2 Sex_i + \hat{\beta}_3 BMI_i + \hat{\beta}_4 I(Day_{ij} = Monday)$$
$$+ \cdots + \hat{\beta}_9 I(Day_{ij} = Saturday) + \hat{\beta}_{10} I(Interval_{ijk} = 9am - 12pm)$$
$$+ \cdots + \hat{\beta}_{14} I(Interval_{ijk} = 9pm - 12am) + \hat{b}_{0i} + \hat{b}_{1ij}$$

The random generation of Poisson-distributed nonwear counts was performed in SAS.

### 3.3.3 Steps for the Evaluation of the Imputation Models

The goal of the evaluation was to see how well the three proceeding models impute the "true" nonwear count $Y_{ijk}^{nw}$ and the "true" overall total count $T_{ijk} = Y_{ijk}^{w} + Y_{ijk}^{nw}$.

**4.** For each replicate data set, we applied each of the three multilevel imputation models, with a random intercept for subject and a random intercept for day, to the wear counts data $Y_{ijk}^{w}$ (randomly generated) and $w_{ijk}$ (observed wear time minutes). All model parameter estimates from the models were retained. The three models are:

(a) Multilevel Poisson with offset (i.e., $\gamma = 1.0$)

(b) Multilevel Poisson with log wear time as a covariate (i.e., estimate $\gamma$)

(c) Multilevel Negative Binomial with offset (i.e., $\gamma = 1.0$)

The Negative Binomial model with log wear time as a covariate was excluded due to convergence issues.

**5.** Using the prediction equation of the fitted model (based on associated parameter estimates including the random effects fixed at their predicted values) perform $m = 1, \dots, 20$ imputations to obtain the nonwear counts $Z_{ijk}^{(m)}$ such that $Z_{ijk}^{(m)} \sim Poisson(\mu_{ijk}^{nw})$ for (a) and (b) and $Z_{ijk}^{(m)} \sim NB(\mu_{ijk}^{nw}, \theta)$ for (c) where $\theta$ is the dispersion parameter for the Negative Binomial distribution. All three models were fit using SAS. The 'rnegbin' function in R 3.4.3 allows the user to use the mean predicted counts and dispersion parameter, from (c) to obtain randomly generated nonwear counts that are distributed as Negative Binomial. Alternatively, counts can be randomly generated in SAS using a combination of the Poisson and Gamma distributions. The nonwear counts randomly generated in R were then imported into SAS for use in the remaining analyses.

**6.** Within a replicate, the imputed nonwear count, $Z_{ijk}^{nw}$, was obtained by taking the mean of the $m =$ 20 imputations.

**7.** Each imputed nonwear count (from each model) was compared to the "true" nonwear count, $Y_{ijk}^{nw}$, by computing percent relative bias: $\left\{\left(\sum_{k=1}^{12,182} Y_{ijk}^{nw} - \sum_{k=1}^{12,182} Z_{ijk}^{nw}\right) / \sum_{k=1}^{12,182} Y_{ijk}^{nw}\right\} \times 100\%$.

**8.** The average total imputed count, $C_{ijk}$, (mean of $C_{ijk}^{(m)}$) was compared to the "true" overall total count, $T_{ijk}$, by computing percent relative bias: $\left\{\left(\sum_{k=1}^{12,182} T_{ijk} - \sum_{k=1}^{12,182} C_{ijk}\right) / \sum_{k=1}^{12,182} T_{ijk}\right\} \times 100\%$.

There was a total of 12,182 intervals that were either entirely missing (frequency = 2,220) or partially observed (frequency = 9,962).

**9.** The last step was to take the average of percent relative bias for imputed nonwear count and average total imputed count over the 200 replicates. The results were reported overall, and by day of the week, and interval of the day.

Steps 2-9 were repeated for the case of $\gamma = 1.0$. Beginning with step 2, the data were generated using the multilevel Poisson mixed model with an offset equaled to one and thus $\gamma = 1.0$.

### 3.4 Results

The analysis sample of n = 1,000 participants made up 7.7% of the overall HCHS/SOL study population (Table 3.1). Approximately 55% of the analysis sample consisted of women and 76% were between the ages of 25-54. Roughly 78% of the analysis sample was classified as overweight or obese. To assess the performance of the multilevel generalized mixed model for imputation, we selected a group of participants with an amount of missing data (i.e., intervals) that warranted attention and some corrective measure. Multiple imputation is commonly employed to remediate the potential bias that arises when the assumption of MCAR is not tenable and the amount of missing data exceeds 5%. There was a total of 35,988 intervals among the 1,000 participants (Table 3.2). Overall, approximately 34% of the intervals were either entirely missing (6%) or partially observed (28%). The remaining 66% of the intervals were fully observed. Measures of percent relative bias for the imputed nonwear count and the average total imputed count are reported in Tables 3.4 and 3.5, respectively.

### 3.4.1 Evaluation of Imputation Models: Counts per interval Generated with $\boldsymbol{\gamma = 1.0}$ and $\boldsymbol{\gamma = 1.3}$

Results for the imputation models where the data was generated using the multilevel Poisson mixed model with $\gamma = 1.0$ and $\gamma = 1.3$ are presented here. In general, a multilevel generalized mixed model with the coefficient of the offset (exposure), $\gamma$, equal to one versus $\gamma > 1.0$ will yield lower predicted counts (Section 2.6, Table 2.9). The percent relative biases for the nonwear counts and average total counts per interval imputed using the Poisson evaluation models were generally smaller than they were for the Negative Binomial evaluation models (Tables 3.4 and 3.5). This can be partially attributed to the specification of the Negative Binomial and Poisson distribution models; the Negative Binomial distribution has one more parameter than the Poisson distribution to adjust the variance independent of the mean.

At the outset we hypothesized that evaluation models that are accurately specified (i.e., coincide with the specification of the data generating model) would produce lower estimates of percent relative bias. The Poisson evaluation models, for imputed nonwear and average total counts and with $\gamma = 1.0$, did not consistently yield the lowest values of percent relative bias as expected. The percent relative biases for the Poisson evaluation model with $\gamma \neq 1.0$ were slightly lower for the 6:00 a.m. – 9:00 a.m. and 9:00 a.m. – 12:00 p.m. intervals and Tuesday, for example (Tables 3.4 and 3.5). However, the Poisson evaluation models, for imputed nonwear and average total counts with $\gamma \neq 1.0$, consistently yielded the lowest values of percent relative bias for data generated with $\gamma = 1.3$. The percent relative biases for the Poisson evaluation models for imputed nonwear counts were similar and generally ranged from 55% to 77% (Table 3.4). In other words, the "true" nonwear counts, $Y_{ijk}^{nw}$, were 55% to 77% larger than the imputed nonwear counts, $Z_{ijk}^{nw}$. Percent relative biases for the Negative Binomial evaluation model ranged from 63% to 82%. The "true" nonwear counts for the Negative Binomial model were 63% to 82% larger than the imputed nonwear counts. The "true" total counts, $T_{ijk}$, were 29% to 51% larger than the imputed total counts, $C_{ijk}$. The percent relative biases for the average total imputed counts were smaller compared to the imputed nonwear counts as the "true" wear counts, $Y_{ijk}^{w}$, were known, which were included in the calculation of the $T_{ijk}$ (i.e., $T_{ijk} = Y_{ijk}^{w} + Y_{ijk}^{nw}$).

For this study sample, in addition to 34% missing data rate, missing data (i.e., nonwear) tended to be highest on Sunday and Saturday and during the 6:00 a.m.-9:00 a.m. and 9:00 p.m.-12:00 midnight intervals (Section 2.6, Figures 2 and 3). Consequently, percent relative biases for imputed nonwear and average total counts for the Poisson and Negative Binomial evaluation models were highest on Sunday and Saturday and during the 6:00 a.m.-9:00 a.m. and 9:00 p.m.-12:00 midnight intervals (Table 3.5). Given the moderate amount of missing data and the possibility that the data were not missing completely at random, the multilevel generalized mixed model may be a conservative approach for imputing missing accelerometer nonwear counts.

### 3.5 Discussion

Measures of imputation bias are useful in determining the performance of an imputation procedure. In this paper, we set out to evaluate the performance of the multilevel generalized mixed model. Accelerometer wear time, the exposure, was an important component for this imputation technique. When creating the data for the simulation, we experimented with treating wear time as an offset, with a coefficient equal to 1, and treating wear time as a covariate with a coefficient equal to 1.3. To evaluate the imputation technique, we specified three multilevel mixed models; Poisson with $\gamma = 1$, Poisson with $\gamma \neq 1$, and Negative Binomial with $\gamma = 1$. Percent relative bias was reported for the imputed nonwear counts per interval and average total imputed counts per interval. Imputation bias tended to increase as the proportion of missing data increased.[9,15]

In chapter 2 (Figure 3), the highest amounts of missing data were observed for Sunday (17.4%) and Saturday (12.1%). The percent relative bias for imputed nonwear counts and imputed average total counts, across all simulation scenarios, was observed for the same days (Tables 3.4 and 3.5). The simulation results for the data generation model with $\gamma = 1.3$ and the evaluation model with $\gamma \neq 1$ yielded the lowest estimates of percent relative bias. However, the results were inconclusive for the data generation and evaluation models with $\gamma = 1.0$.

There were a couple of limitations to this simulation study. First, 20 imputations may have been insufficient given the hierarchical structure of the data. Some investigators recommend 20 imputations for single-level models.[56] The number of imputations for multilevel imputation has not yet been systematically examined and because increasing the number of imputations generally leads to an increase in power for subsequent analyses, a larger number of imputations may be advisable (e.g., 50 or 100).[57] A drawback of a large number of imputations for multilevel data is an increase in computational resources.[57] Second, convergence issues may arise with certain data generation and imputation evaluation models. For example, convergence issues were observed for the Negative Binomial distribution, with $\gamma \neq 1$, when used for data generation and imputation model evaluation. This may have been attributed to the dispersion parameter that has to be estimated with Negative Binomial models.

Based on the mixed results, we were unable to find any meaningful evidence in support of our hypothesis, which was that percent relative bias would be smaller for MGMM imputation evaluation models that are concordant with MGMM data generation models. One explanation for this may be that the imputation evaluation models may be misspecified in certain scenarios. For example, in a secondary analysis, Poisson marginal evaluation models fit to the multilevel data yielded percent relative biases that were approximately 10% lower than the corresponding percent relative biases for the multilevel Poisson mixed evaluation models (Tables 3.4 and 3.5). Thus, the Poisson marginal evaluation model was the "best" model. Too many sources of random variability (i.e., steps 3i and 3ii) may have also contributed to the inconclusive results. Even though the marginal Poisson evaluation model was the best model, percent relative biases for all data generation and imputation model evaluations exceeded 10%, which was unsatisfactory. Thus, we would not recommend the MGMM imputation technique based on the current simulation study.

In future simulation studies, one modification to consider is to draw the random effects from their predictive posterior distributions instead of from a normal distribution with a mean of zero and variance $\sigma^2$. Additional modifications to the simulation should impose different amounts of missing data that are generated with different missing mechanisms. A variety of evaluation models should also be implemented.

## 3.6 Tables and Figures

Table 3.1. Weighted descriptive characteristics for the full study population, n=16,415, and the analysis subset, n=1,000, HCHS/SOL 2008-2011

| Characteristic | Total n | Weighted Percent | Total n | Weighted Percent |
|---|---|---|---|---|
| Overall | 16,415 | 100.0 | 1,000 | 7.7 |
| Gender: | | | | |
| Female | 9,835 | 52.1 | 605 | 54.6 |
| Male | 6,580 | 47.9 | 395 | 45.4 |
| Age (yr): | | | | |
| 18-24 | 1,665 | 16.8 | 68 | 12.5 |
| 25-34 | 2,082 | 21.8 | 86 | 16.2 |
| 35-44 | 2,954 | 21.2 | 171 | 21.3 |
| 45-54 | 4,922 | 18.9 | 319 | 20.8 |
| 55-64 | 3,460 | 12.8 | 251 | 18.0 |
| 65+ | 1,332 | 8.5 | 105 | 11.3 |
| Body mass index (kg/m$^2$): | | | | |
| Underweight | 130 | 1.2 | 8 | 0.8 |
| Normal | 3,191 | 22.1 | 192 | 20.8 |
| Overweight | 6,116 | 37.2 | 383 | 39.3 |
| Obese | 6,907 | 39.6 | 414 | 39.1 |

Table 3.2. Frequency and percent of entirely missing, partially missing, and fully observed intervals for the analysis subset, n=1,000

| Interval Missing Status | Frequency | Percent | Cumulative Percent |
|---|---|---|---|
| Entirely missing | 2,220 | 6.2 | 6.2 |
| Partially missing | 9,962 | 27.6 | 33.9 |
| Fully observed | 23,806 | 66.2 | 100.0 |
| Total | 35,988 | 100 | 100.0 |

Table 3.3. Imputation model parameter estimates and standard errors for total counts per interval based on complete cases using a multilevel generalized linear mixed model (MGMM) with offset ($\gamma = 1$) and with log-wear time as covariate ($\gamma = 1.046$), n=1000 participants (35,988 intervals/observations)

| | MGMM $\gamma = 1$ | MGMM $\gamma = 1.046$ |
|---|---|---|
| Variable | Estimate (SE) | Estimate (SE) |
| $\log(N_{ijt})$ | 1.000 (--) | 1.046 (0.049) |
| Intercept | 5.960 (0.136) | 5.049 (0.259) |
| Age (yr) | -0.015 (0.002) | -0.005 (0.002) |
| Male | 0.267 (0.042) | 0.266 (0.042) |
| BMI (kg/m$^2$) | -0.021 (0.004) | -0.015 (0.004) |
| Day of Week | | |
| Monday | 0.179 (0.028) | 0.176 (0.029) |
| Tuesday | 0.169 (0.027) | 0.166 (0.028) |
| Wednesday | 0.160 (0.029) | 0.156 (0.031) |
| Thursday | 0.206 (0.027) | 0.202 (0.028) |
| Friday | 0.238 (0.026) | 0.233 (0.027) |
| Saturday | 0.114 (0.027) | 0.112 (0.027) |
| Interval | | |
| 9:00am-12:00pm | 0.165 (0.040) | 0.159 (0.043) |
| 12:00pm-3:00pm | 0.183 (0.048) | 0.176 (0.052) |
| 3:00pm-6:00pm | 0.073 (0.049) | 0.066 (0.053) |
| 6:00pm-9:00pm | -0.143 (0.049) | -0.149 (0.053) |
| 9:00pm-12:00am | -0.604 (0.053) | -0.603 (0.053) |
| Var($b_{0i}$) = $\sigma_0^2$ | 0.886 (0.409) | 0.883 (0.431) |
| Var($b_{1ij}$) = $\sigma_1^2$ | 0.316 (0.023) | 0.314 (0.022) |

MGMM = multilevel generalized mixed model.

Table 3.4. Percent relative bias for the evaluation of five models used to impute **nonwear counts per interval**: Multilevel Poisson and Negative Binomial mixed models each with $\gamma = 1$ and $\gamma \neq 1$ (the coefficient of log-wear time minutes) where counts per interval were initially generated using a multilevel Poisson mixed model with $\gamma = 1.0$ and $\gamma = 1.3$.

| | $\gamma = 1.0$ | | | | | $\gamma = 1.3$ | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Characteristic | Poisson $\gamma = 1$ | Poisson $\gamma \neq 1$ | NB $\gamma = 1$ | GEE, Poisson $\gamma = 1$ | GEE, Poisson $\gamma \neq 1$ | Poisson $\gamma = 1$ | Poisson $\gamma \neq 1$ | NB $\gamma = 1$ | GEE, Poisson $\gamma = 1$ | GEE, Poisson $\gamma \neq 1$ |
| Overall | 68.5 | 68.4 | 74.5 | 58.2 | 58.5 | 67.3 | 66.0 | 76.0 | 55.5 | 56.1 |
| Interval | | | | | | | | | | |
| 6:00am-9:00am | 76.6 | 76.4 | 81.1 | 67.7 | 67.8 | 76.1 | 75.2 | 82.2 | 66.2 | 66.9 |
| 9:00am-12:00pm | 71.0 | 70.2 | 76.9 | 59.9 | 60.2 | 67.9 | 66.7 | 76.7 | 55.8 | 55.9 |
| 12:00pm-3:00pm | 68.0 | 68.4 | 74.7 | 61.0 | 61.4 | 66.3 | 64.5 | 75.7 | 58.1 | 58.3 |
| 3:00pm-6:00pm | 66.8 | 66.1 | 72.9 | 58.9 | 59.6 | 64.7 | 62.8 | 74.4 | 55.3 | 55.5 |
| 6:00pm-9:00pm | 60.9 | 61.4 | 67.6 | 50.9 | 51.4 | 58.7 | 57.5 | 70.5 | 46.4 | 46.5 |
| 9:00pm-12:00am | 55.6 | 56.7 | 63.6 | 41.2 | 41.6 | 55.7 | 54.0 | 67.4 | 38.5 | 39.7 |
| Day of Week | | | | | | | | | | |
| Sunday | 68.6 | 68.9 | 74.5 | 58.9 | 59.0 | 68.7 | 67.3 | 76.6 | 57.4 | 57.4 |
| Monday | 66.8 | 67.1 | 72.9 | 57.7 | 57.9 | 65.9 | 64.8 | 75.2 | 54.9 | 55.3 |
| Tuesday | 69.1 | 68.5 | 75.3 | 56.8 | 57.3 | 67.2 | 65.9 | 76.5 | 53.9 | 55.1 |
| Wednesday | 64.5 | 64.4 | 71.4 | 55.6 | 55.9 | 62.9 | 61.5 | 72.2 | 51.8 | 52.7 |
| Thursday | 69.9 | 69.9 | 75.1 | 58.4 | 58.9 | 67.8 | 66.8 | 76.1 | 54.9 | 55.2 |
| Friday | 69.2 | 69.0 | 75.4 | 59.5 | 59.5 | 68.1 | 66.1 | 76.4 | 55.9 | 56.6 |
| Saturday | 70.4 | 70.0 | 75.8 | 59.3 | 59.6 | 68.7 | 67.9 | 77.8 | 56.8 | 58.1 |

NB = Negative Binomial, GEE = generalized estimating equations. Initial counts per interval were generated using a multilevel generalized Poisson mixed model with log-wear time as a covariate (i.e., coefficient $\gamma = 1.3$) and log-wear time as an offset (i.e., $\gamma = 1.0$). The five evaluation models were applied to each of the two initial data sets to see how well each model imputes nonwear counts per interval.

Table 3.5. Percent relative bias for the evaluation of five models used to impute **average total counts per interval**: Multilevel Poisson and Negative Binomial mixed models each with $\gamma = 1$ and $\gamma \neq 1$ (the coefficient of log-wear time minutes) where counts per interval were initially generated using a multilevel Poisson mixed model with $\gamma = 1.0$ and $\gamma = 1.3$.

| | $\gamma = 1.0$ | | | | | $\gamma = 1.3$ | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Characteristic | Poisson $\gamma = 1$ | Poisson $\gamma \neq 1$ | NB $\gamma = 1$ | GEE, Poisson $\gamma = 1$ | GEE, Poisson $\gamma \neq 1$ | Poisson $\gamma = 1$ | Poisson $\gamma \neq 1$ | NB $\gamma = 1$ | GEE, Poisson $\gamma = 1$ | GEE, Poisson $\gamma \neq 1$ |
| Overall | 38.0 | 38.0 | 41.7 | 32.5 | 32.6 | 37.9 | 36.9 | 42.8 | 31.1 | 31.3 |
| Interval | | | | | | | | | | |
| 6:00am-9:00am | 46.4 | 45.9 | 49.7 | 41.4 | 41.4 | 47.1 | 46.5 | 50.7 | 41.1 | 41.3 |
| 9:00am-12:00pm | 35.9 | 34.9 | 38.9 | 30.1 | 30.2 | 33.9 | 32.9 | 38.0 | 27.5 | 27.4 |
| 12:00pm-3:00pm | 35.3 | 36.1 | 39.1 | 32.1 | 32.1 | 34.7 | 33.5 | 39.5 | 30.3 | 30.3 |
| 3:00pm-6:00pm | 33.5 | 33.7 | 36.6 | 29.7 | 30.0 | 32.5 | 31.1 | 37.6 | 27.5 | 27.4 |
| 6:00pm-9:00pm | 31.1 | 31.6 | 35.1 | 26.2 | 26.5 | 29.8 | 29.2 | 36.2 | 23.5 | 23.4 |
| 9:00pm-12:00am | 31.8 | 32.8 | 36.4 | 23.6 | 23.8 | 32.5 | 31.5 | 39.7 | 22.4 | 23.1 |
| Day of Week | | | | | | | | | | |
| Sunday | 42.6 | 42.6 | 45.9 | 36.4 | 36.4 | 43.3 | 42.2 | 47.9 | 36.0 | 35.7 |
| Monday | 36.2 | 36.9 | 40.3 | 31.6 | 31.8 | 36.6 | 35.9 | 42.2 | 30.5 | 30.5 |
| Tuesday | 38.0 | 37.5 | 41.7 | 31.3 | 31.6 | 37.0 | 36.3 | 42.3 | 29.7 | 30.5 |
| Wednesday | 34.1 | 34.2 | 38.7 | 30.0 | 30.0 | 34.0 | 33.0 | 38.7 | 27.9 | 28.5 |
| Thursday | 38.1 | 37.5 | 40.6 | 31.7 | 31.9 | 36.7 | 35.9 | 41.1 | 29.6 | 29.3 |
| Friday | 36.0 | 36.0 | 39.7 | 31.4 | 31.3 | 36.2 | 34.7 | 40.8 | 29.4 | 29.6 |
| Saturday | 39.9 | 39.9 | 43.5 | 33.8 | 34.0 | 39.4 | 38.9 | 44.5 | 32.6 | 33.4 |

NB = Negative Binomial, GEE = generalized estimating equations. Initial counts per interval were generated using a multilevel generalized Poisson mixed model with log-wear time as a covariate (i.e., coefficient $\gamma = 1.3$) and log-wear time as an offset (i.e., $\gamma = 1.0$). The five evaluation models were applied to each of the two initial data sets to see how well each model imputes average total counts per interval.

# CHAPTER 4: APPLICATION OF THE MULTILEVEL GENERALIZED MIXED MODEL AND OTHER APPROACHES THAT ADDRESS ACCELEROMETER MISSING DATA: THE HCHS/SOL YOUTH

## 4.1 Introduction

A physically active lifestyle among children promotes short- and long-term physical and mental well-being. Regular physical activity in children and adolescents lowers their risk of chronic diseases and improves their chances of becoming healthy adults.[39] Accelerometry is commonly used to measure physical activity in large epidemiologic studies. Strengths of accelerometry include the accuracy with which the device is able to capture physical activity and the ability to record large amounts of data. One limitation of hip-worn devices is that they do not capture upper-body movement well.[6] They primarily measure locomotor activity.[6] Lastly, accurately and precisely summarizing accelerometer data in the presence of missing data creates a variety of analytical challenges. The challenges of analyzing counts data is only applicable to certain accelerometers that apply a proprietary conversion algorithm to the raw data.

Devices, such as the Phillips Respironics Actical accelerometer, worn on the hip, can be used to estimate physical activity and sedentary behavior. These devices measure acceleration in units of gravity ($g \approx 9.81$ m/s$^2$) at a specific frequency (e.g., 30 times/s or 30 Hz).[10] The company aggregates accelerations to activity counts per user-specified unit of time, known as an epoch.[10] Nonwear time, resulting in a period of consecutive zero counts, later becomes missing data. This subsequent missing data is a result of data aggregation methods employed by investigators to reduce the bias that may arise when all observed data is used to compute summary statistics. For example, an estimate for total counts for a particular day that is based on all observed data will underestimate the true level of physical activity if the device is worn for only part of the day.[9] To circumvent this potential bias, researchers use the inclusion criteria that participants have a minimum number of days each with a minimum number of hours of wear (e.g., ≥ 3 d with ≥ 8 h wear time/d) in order to be included in the final analysis sample. Participants with accelerometer data that do not meet the aforementioned criteria are commonly categorized as *missing* and/or non-adherent, and subsequently excluded from analyses.[9,12]

The concern with this approach is that it assumes physical activity and sedentary behavior are missing completely at random (MCAR) during nonwear. For example, the MCAR assumption would be violated if participants are more (or less) likely to wear the device during nonwear.[9] While directly addressing accelerometer missing data via some imputation method is less conventional, a number of researchers have had success in obtaining more reliable estimates of physical activity and sedentary behavior using various statistical methods.[9,18,19,21,22,24]

Several authors have proposed procedures to impute missing accelerometer data. Catellier et al. implemented a single value imputation method, the expectation-maximization algorithm, and multiple imputation (MI) to impute missing accelerometer data for the entire day and interval of the day.[9] This study showed that the performance of each imputation method depends on the proportion of missing data, the correlation of activity across days of the week, and the missing data mechanism.[9] Using all available data from adherent and non-adherent days, Lee and Liu performed MI of daily accelerometer counts/min and missing steps per 60-s epoch, respectively, using additive regression, bootstrapping, and predictive mean matching.[18,19] In addition to using data from adherent and non-adherent days, Xu proposed a mixed model technique for imputation, which accounts for variation between and within participants.[21] Some advanced techniques for imputing missing accelerometer data include Bayesian and zero-inflated models.[22,24]

In Chapter 2, we presented a statistical method to impute missing accelerometry that incorporates the use of all available data, variability from within and between participants, and is well suited to account for multivariate count data under a complex survey design: the multilevel generalized mixed model (MGMM). We concluded that failing to include wear time as a covariate in the imputation model may lead to an underestimation of average counts/min and total counts per interval (Tables 2.3, 2.9). This was indicated by lower imputed values based on the models with $\gamma = 1$ versus $\gamma$ =1.083. We also concluded that the rate of counts/min may not be the same during wear and nonwear periods.

In Chapter 3, we presented a simulation study to assess the performance of the MGMM imputation method by using a measure of percent relative bias. The simulation results for the data generation model with $\gamma = 1.3$ and the evaluation model with $\gamma \neq 1$ yielded the lowest estimates of percent relative bias. However, the results were inconclusive for the data generation and evaluation models with $\gamma = 1.0$. For data generated using the MGMM, with $\gamma = 1.3$ and $\gamma \neq 1$, the Poisson regression evaluation marginal models (e.g., without random effects) yielded the lowest estimates of percent relative bias compared to all of the multilevel generalized mixed evaluation models. We were not able to recommend the MGMM method for imputation based on the simulation results alone as the percent relative biases were greater than 10% and thus, unsatisfactory.

Accelerometer data from The Hispanic Community Health Study/Study of Latinos (HCHS/SOL), 2008 – 2011, an adult population, was used in the two previous chapters. Using data from The Hispanic Community Children's Health Study/Study of Latino Youth (SOL Youth), the purpose of this paper was to compare the MGMM imputation method with five different approaches used to address missing accelerometer data. These comparisons were made using descriptive statistics (e.g., mean and standard deviation) for average counts/min as well as linear regression model parameter estimates with average counts/min as a covariate and body mass index (BMI) (kg/$m^2$) as the outcome. The motivation for the latter is explained below.

Most of the studies on missing accelerometry approaches do not investigate how the proposed technique performs in models with physical activity as an independent variable.[9,12,22,48] The methods in the literature tend to focus on physical activity as a dependent variable. Thus, we investigated whether or not imputing data at the interval, day, or participant level makes a difference on point estimates and tests for associations between average counts/min and BMI (kg/$m^2$). We hypothesize that there is a significant association between average count/min and BMI when missing values are imputed at the interval level using the multilevel generalized mixed model.

## 4.2 Physical Activity Demonstration Example

Accelerometer data from The Hispanic Community Health Study/Study of Latino Youth (SOL Youth), 2012 – 2014, was used to compare six methods used for handling missing accelerometer data.

### 4.2.1 Study Population

The Hispanic Community Children's Health Study/Study of Latino Youth is a multicenter community-based study of Hispanic/Latino children, aged 8 – 16, living in the United States whose parents/legal guardians participated in the Hispanic Community Health Study/Study of Latinos (HCHS/SOL), a large community-based cohort study of Hispanic/Latino adults living in the United States.[58] The SOL Youth Study was designed to examine associations of youth's lifestyle behaviors and cardiometabolic risk factors with (1) youth's acculturation and parent-child differences in acculturation; (2) parenting strategies, family behaviors, and parental health behaviors; and (3) youth's psychosocial functioning.[58] Between 2012 and 2014, 1,600 children in four communities (the Bronx, New York; Chicago, Illinois; Miami, Florida; San Diego, California) were selected from a population-based sample of Hispanic households whose adult members are enrolled in the

HCHS/SOL where recruitment was implemented through a two-stage area household probability design.[45,46,58]

### 4.2.2 Data Measures

Accelerometer data was obtained from participants at the four HCHS/SOL field centers. During the baseline clinic visit, participants were asked to wear the Actical™ accelerometer (model 198-0200-03, Minimiter Respironics®, Bend, OR) for seven days during waking hours. They were instructed to undertake usual activities while wearing the monitor on the hip and to remove it only for swimming, bathing, and sleeping. In addition to receiving instructions on proper wear during the clinic visit, participants were given written instructions and a phone number to call if questions arose during the 7-day monitoring period. At the end of the monitoring period, participants returned the Actical to the field center in person or via mail.

The Actical accelerometer measures movement in all directions and was programmed to record counts and steps in 15-second epochs.[59] Average counts/min was the primary measure for this study. Nonwear time was defined as at least 90 minutes of consecutive zero counts, with the allowance for intervals of up to two minutes of nonzero counts if the 30 minutes preceding and following these intervals were consecutive zero counts.[48]

### 4.2.3 Analytic Sample

A total of 1,466 children participated in the SOL Youth Study. Of the 1,466 participants, 1,238 wore the accelerometer and 1,104 were classified as adherent (i.e., ≥8 h of wear time for ≥3 d). Our interest was in estimating physical activity when participants have been instructed to wear the device (i.e., during waking hours). Thus, the start of our monitoring day at 6:00 a.m. Moreover, we wanted to be consistent with the accelerometer analytic sample for the HCHS/SOL, thus only data recorded between 6:00 a.m. and 12:00 a.m. each day (18 h) were included. As expected, nonwear was highest between 12:01 a.m. and 5:59 a.m. when children were sleeping.

### 4.2.4 Physical Activity Outcome

Accelerometer data is normally aggregated at the day or participant level in child studies. This level of reporting is consistent with the U.S. Department of Health and Human Services' reporting of recommendations for physical activity.[1,35] The outcome of interest for this study was average counts/min per participant which is a measure of total volume of physical activity. There was a total of 4,320 15-second epochs per day between 6:00 a.m. and 12:00 a.m. Counts for every 4 consecutive epochs were added to obtain counts/min within a day. Similarly, counts for every 240 consecutive epochs were added to obtain counts/h within a day. For an individual, total counts on a given day were summed and then divided by the total accelerometer wear time, in minutes, for that day to obtain average counts/min/day. Next, average counts/min/day for a participant were averaged across their observed monitoring days to obtain that participant's measure of average counts/min.

For the MGMM imputation approach, data were imputed at the interval level then aggregated at the participant level. Specifically, each 18-h monitoring day was divided into six intervals of equal length: (1) 6:00 a.m.-9:00 a.m., (2) 9:00 a.m.-12:00 p.m., (3) 12:00 p.m.-3:00 p.m., (4) 3:00 p.m.-6:00 p.m., (5) 6:00 p.m.-9:00 p.m., and (6) 9:00 p.m.-12:00 a.m. The maximum possible wear time for each interval was three hours (180 min).

**Covariates**

Demographic factors self-reported during the baseline exam include: age, sex, Hispanic/Latino group, parent's education level, language preference, and immigration status.[59] Using standardized protocols, weight was measured for youth on a digital scale (Tanita Body Composition Analyzer, TBF 300, Japan) to the nearest 0.1 kg and height to the nearest cm using a wall-mounted stadiometer (SECA 222, Germany). Body mass index was calculated as weight in kilograms divided by height in meters squared. BMI was also represented as age- and sex-standardized BMI percentiles.[59]

### 4.2.5 Statistical Analyses

The statistical approaches for analyzing incomplete accelerometer data are described in this section. The analysis sample used to demonstrate each approach is also described. The term "approach" in this context refers to how missing data is handled. One approach is to do nothing to remediate the missing data and to analyze the data as is (e.g., Actical). Another approach, considered to be ad hoc, is to analyze data only for participants with a sufficient amount of data (e.g., Adherent). The last type of approach involves the implementation of a weighting or imputation technique to remediate the missing data (e.g., IPW, MI, MGMM). Average counts/min were summarized for each approach overall and by sex, age group, and BMI ($kg/m^2$) group. The BMI groups were constructed according to the Center for Disease Control and Prevention's growth chart (as of March 26, 2014).[59,60] Using BMI percentiles, participants were classified as (1) underweight (BMI <5$^{th}$ percentile), (2) normal weight (BMI 5-84$^{th}$ percentile) (3) overweight (BMI 85-94$^{th}$ percentile), (4) obesity (BMI 95$^{th}$+ percentile, BMI <35 and 125% of 95$^{th}$ percentile), and (5) severe obesity (BMI≥35 or 125% of 95$^{th}$ percentile).[60] With regard to approaches 4-6, relative efficiency (RE), in units of variance, was compared for 5 and 10 imputations to decide on the optimal number of imputations. Relative efficiency is approximately a function of the proportion of missing data, $\tau$, and the number of imputations, *m*.[30,29]

$$RE = \left(1 + \frac{\tau}{m}\right)^{-1}$$

**Approach 1: Actical**

There was a total of 1,238 participants that wore the Actical for an average of 11.4 hours per day. Everyone in this group wore the device for at least three days. No imputation of average counts/min was performed for the Actical group. All descriptive information and regression analyses for this group were performed using all available data.

**Approach 2: Adherent**

Adherent participants, N = 1,104, wore the accelerometer for an average 12.3 hours per day. Everyone in this group wore the device for at least 6 days. There was no imputation of average counts/min performed for the adherent group of participants. All descriptive information and regression analyses for this group were performed using all available data.

**Approach 3: Inverse probability weighting (IPW)**

One approach for handling missing accelerometer data that does not require any assumptions regarding the conditional distribution of the missing data given the observed data is inverse probability weighting (IPW).[23] Instead of "filling in" missing values with plausible values as is done with imputation, an adjustment to the analysis is made by weighting the observed data (e.g., Adherent participants) appropriately which reduces the bias in estimates that is induced when a complete case analysis (e.g., all Actical participants) is performed and the data are not missing completely at random.[23] With IPW, more influence is given to cases that have a small probability of being cases and that are representative of non-cases with missing data.[34,61] Participants classified as cases are given an inflated weight to compensate for missing data on participants with similar outcome profiles but whose data is missing.[34,61]

Cases for this study were defined as adherent participants (i.e., ≥8 h of wear time for ≥3 d). The first step for calculating the IPW was to fit a logistic regression model to obtain the predicted probability of being adherent in the full sample population (N = 1,466). The outcome variable was dichotomous and equaled one if the participant was adherent and zero otherwise. Next, the predicted probability of being adherent was used to compute an inverse probability weight for each of the N = 1,104 participants. Lastly, the sampling weight was combined with the IPW by multiplying the two to obtain a single weight for use in subsequent analyses involving objectively measured physical activity.[34] All covariates plus all pairwise interactions for the logistic regression model, except for STRATA and IMGEN which were used as main effects, are listed in Table 4.2. All IPW analyses were conducted using SAS software 9.4 (SAS Institute, Cary, NC). The application of the IPW in a regression analysis will be described later in section 4.3.6; "Analysis with Complete Counts".

**Approach 4: Multiple imputation (MI) using the fully conditional specification (FCS)**

In contrast to the IPW approach, MI requires a model for the distribution of missing data given the observed data.[34] Multiple imputation using the FCS approach was used to impute missing accelerometer average counts/min at the participant level in SAS software (version 9.4). Missing data for covariates and the response variable were imputed for all 1,466 participants which included 228 participants without any accelerometer data at all. The MI procedure using the FCS method in SAS is not capable of analyzing outcome variables that are distributed as Poisson or Negative Binomial and thus average counts/min were analyzed as continuous. In general, the FCS approach employs an iterative algorithm to impute missing values. All variables involved in the imputation procedure, outcome and covariates, were listed in order of increasing missing data amount in the VAR statement. For this study, the order of variables in the VAR statement was: Age, BMI, Sex, Strata, Sample weight, then Average counts/min. The order of variables determined the pattern of missingness in the imputed data set.[27]

Each iteration of the FCS algorithm moves through this sequence of variables, one-by-one, performing two steps (Posterior-step and Imputation-step) at each iteration for each variable.[27] During the P-step, the current values of the observed and imputed values for the imputation model variables are used to derive the predictive distribution of the missing values for the target variable.[27] The estimates created at each iteration for each variable are then used in the next I-step. The

Imputation-step simulates the missing values for each observation independently.[27] Essentially, a univariate model for each variable in the imputation was fit. The univariate models for the continuous variables: Age, BMI, Sample weight, and Average counts/min, were fit using the keyword REG in the FCS procedure. The dichotomous variable Sex was fit using the LOGISTIC keyword. The nominal categorical variable Strata was fit using the DISCRIM keyword. Five imputations were performed.

**Approach 5: Lee**

Lee developed an approach to impute missing accelerometer data that uses information from adherent (i.e., ≥ 8 h wear per day) and non-adherent days (i.e., < 8 h wear per day).[19] The ultimate goal of Lee's approach is to impute average counts/min (per day) for all non-adherent days such that each participant has seven days of data each with at least eight hours of wear time.[19] This approach uses accelerometer data from adherent (≥ 8 h) and non-adherent (< 8 h) days for adherent and non-adherent participants, thus N = 1,238.[19] Participants with no accelerometer data were excluded (N = 228).

Average counts per minute and wear time were imputed separately. Wear time was imputed with the assumption that all participants should have worn the accelerometer at least eight hours on all days.[19] Accelerometer counts,$w_{ij}$, observed during wear time, $w_{ij}$, for participant *i* on day *j*, were combined with imputed counts in the remaining $8 - w_{ij}$ hours to obtain the final imputation, $y_{ij}^{imp}$.[19] Five imputations were performed using additive regression (AR), bootstrapping, and predictive mean matching, available in the Hmisc package, in R (version 3.4.3).[19] Age and BMI (kg/m$^2$), both continuous, and sex were used as predictors in the AR model. The details of the imputation algorithm are as follows:

1) Imputation of wear time. Estimate the mean $\mu$ by using the imputed expected number of adherent hours from AR and bootstrapping. If the imputed expected number of adherent hours is <8 h, it will be designated as 8 h.

2) Imputation of counts per minute. Re-estimate the mean $\eta$ by using imputed counts per minute from AR and bootstrapping.

3) Combine. Combined imputed counts per minute are equal to

$$\left[Imputed\frac{counts}{min} \times (imputed\ expected\ number\ of\ valid\ min - observed\ number\ of\ valid\ min) + observed\ total\ counts\right] \div imputed\ expected\ number\ of\ valid\ minutes$$ [19]

If all seven days were non-adherent for a participant then missing average counts/min were imputed by using mean values adjusted for age, sex, and BMI (kg/m$^2$) from non-missing data.[19] This imputation approach does not account for the complex survey design. Complete datasets were imported into and analyzed using SAS.

**Approach 6: Multilevel Generalized Mixed Model (MGMM)**

The multilevel generalized mixed model imputation approach takes advantage of the richness in information provided by all available data among those with at least eight hours of wear time for at least three days. Similar to Lee's approach, the MGMM approach, for both $\gamma = 1$ and $\gamma \neq 1$, used accelerometer data from adherent ($\geq$ 8 h) and non-adherent (< 8 h) days for adherent and non-adherent participants, thus N = 1,238. Under the assumption of missing at random (MAR), parameter estimates from the MGMM of counts per interval as a function of wear time, age (years), BMI (kg/m$^2$), sex, day (Sunday, Monday,…, Saturday), interval of the day (i.e., (1) 6:00 a.m.-9:00 a.m., (2) 9:00 a.m.-12:00 p.m., (3) 12:00 p.m.-3:00 p.m., (4) 3:00 p.m.-6:00 p.m., (5) 6:00 p.m.-9:00

p.m., and (6) 9:00 p.m.-12:00 midnight) and survey sampling strata were used to obtain unknown counts per interval during nonwear. A 20-level strata variable was used in the imputation model to account for the HCHS/SOL Youth survey design. As the first step of the MI technique, the model parameter estimates were used to fill in five missing values in order to create five complete data sets. Of the 50,589 intervals, 4,450 (9%) were missing. Counts per interval were distributed as Negative Binomial (NB). The variation in the data was in excess of that expected under the Poisson assumption. Age (years), BMI ($kg/m^2$), and wear time minutes were continuous; sex, day of the week, interval, and stratum were categorical. Two MGMM models were explored; one with log-wear time treated as an offset where the coefficient, $\gamma$, equals one and the second with log-wear time as a predictor with a coefficient, $\gamma$, not equal to one. Intervals were nested within day and day within subject, hence the multilevel structure.

Using the model parameter estimates from the MGMM based on wear time, predicted values based on nonwear time were calculated then added to observed values to obtain imputed values for intervals that were partially observed or completely missing. "Completely missing" intervals were determined to be "missing" based on the nonwear definition used in the HCHS/SOL Youth. Nonwear time for each interval was calculated as 180 minutes minus wear time for that interval. Counts for intervals that were completely observed were preserved (i.e., no imputation was performed).

The following equation was used to model counts per interval as a function of the observed wear time and aforementioned covariates.

MGMM:

$$log\left(E\left(\frac{Y_{ijk}}{W_{ijk}^{\gamma}}\middle|\boldsymbol{b_i}\right)\right) = log\left(\frac{\mu_{ijk}}{W_{ijk}^{\gamma}}\right);$$

$$\begin{aligned} log(\mu_{ijk}) = \gamma log(W_{ijk}) + \beta_0 + \beta_1 Age_i + \beta_2 Sex_i + \beta_3 BMI_i \\ +\beta_4 I(Day_{ij} = Monday) + \cdots + \beta_9 I(Day_{ij} = Saturday) \\ +\beta_{10} I(Interval_{ijk} = 9am - 12pm) + \cdots + \beta_{14} I(Interval_{ijk} = 9pm - 12am) \\ +\beta_{15} I(Stratum_{ijk} = 1) + \cdots + \beta_{34} I(Stratum_{ijk} = 19) + b_{0i} + b_{1ij} \end{aligned} \quad (1)$$

Distributed as Negative Binomial, let $Y_{ijk}$ be the observed accelerometer counts for participant *i*, on day *j*, in interval *t*, with mean counts given the random effects $b_i$ given as $E(Y_{ijk}|\boldsymbol{b_i}) = \mu_{ijk}$; *i* = 1,2,3,…, 1,000, *j* = 1,…, 7, and *k* = 1,2…, 6. Let $b_{0i} \sim N(0, \sigma_0^2)$ be the random intercept for participant *i* and $b_{1ij} \sim N(0, \sigma_1^2)$ be the random intercept for day *j* where interval *k* is nested within day *j* and day *j* is nested within participant *i*. The exposure, $W_{ijk}$, is the observed wear time minutes for each interval within each day for each participant. Interpret $\frac{\mu_{ijk}}{W_{ijk}}$ as "the expected value of counts per minute per interval given the random effects $\boldsymbol{b_i}$". The regression coefficient for the exposure is $\gamma$ and $\beta_0 - \beta_{34}$ are the regression coeffients for the remaining model covariates. There were 20 strata. Sex = female, day = Sunday, and interval = 6:00 a.m. – 9:00 a.m. were reference levels. The following steps were undertaken:

**i.** Fit MGMM to wear data.

Fit a multilevel generalized mixed model (MGMM) with observed counts, $Y_{ijk}$, as the dependent variable and log-wear time per interval, age (years), sex, BMI (kg/m$^2$), and day (Sunday, Monday,…, Saturday) as independent variables.

**ii.** To obtain predicted counts for intervals that were completely missing or partially observed, first substitute nonwear time, $W_{ijk}^{nw} = 180 - w_{ijk}$, for $W_{ijk}$ in equation (1). Random effects were estimated using best linear unbiased prediction (BLUP), also known as empirical Bayes estimation.

For intervals that were defined as nonwear and thus contained zero total counts, $W_{ijk}^{nw}$ = 180 minutes. Since participants were assumed to be wearing the accelerometer during the day, in accordance with the SOL Youth protocol, $W_{ijk}$ corresponding to the "missing" intervals were treated as wear time in the imputation model.

**iii.** Impute five values for completely missing intervals where $Y_{ijk}^{imp}$ is imputed by random draw from the Negative Binomial distribution with mean $\hat{\mu}_{ijk}$. In this scenario, $\hat{\mu}_{ijk}$ is the predicted value of mean counts given the random effects $\widehat{\boldsymbol{b}}_{\boldsymbol{i}}$ and $W_{ijk}$ = 180 minutes.

**iv.** Impute five values for partially missing intervals where predicted counts are calculated only for the missing portion of the interval, $Y_{ijk}^{mis}$. Similar to what was done in **iii**, $Y_{ijk}^{mis}$ is imputed by random draw from the Negative Binomial distribution with mean $\hat{\mu}_{ijk}$. In this scenario, $W_{ijk}^{nw} = 180 - w_{ijk}$. The final imputed value, $Y_{ijt}^{imp}$, is the sum of the observed counts, $Y_{ijk}$, and the predicted counts $Y_{ijk}^{mis}$.

### 4.2.6 Analysis with Complete Counts

The data from approaches 1-6 were analyzed using the desired SAS or SUDAAN procedure (i.e., SURVEYMEANS, LOGLINK, or SURVEYREG). Results from multiply imputed data (approaches 4-6) were summarized then combined using PROC MIANALYZE.

#### Descriptive Statistics: Mean and Standard Deviation

For each approach, the mean and standard deviation for average counts/min per person were reported overall and by sex, age group, and BMI (kg/m$^2$) group.

#### Survey Regression: Average Counts/min per Person as the Outcome

Complex survey regression was performed on the data from each approach with average total counts per person as the response. The approach of GEE with a log link function and average wear time as an offset, where appropriate, was used for every approach except MI. Missing average counts per minute were imputed as a continuous variable using MI and thus the complete data were analyzed as continuous using SURVEYREG. The GEE was fit using the LOGLINK procedure in SUDAAN (version 11) with the robust variance estimator of Zeger and Liang (1986) in order to account for the complex survey design and to obtain valid standard errors for model parameter estimates.[53] After sorting the data sets by subject id, the strata and cluster variable for primary sampling units were used on the NEST statement. An exchangeable working correlation structure was estimated. Results of the survey regression model using data from each approach are presented.

**Linear Survey Regression: Association between Average Counts/min and BMI (kg/m$^2$)**

In many accelerometer studies, researchers are interested in how measures of physical activity may affect health outcomes of interest. To investigate how average counts/min, after applying 6 approaches for dealing with missing data, impact BMI (kg/m$^2$), we carried out a linear regression analysis with BMI (kg/m$^2$) as the outcome and average counts/min, age, sex, Hispanic/Latino background, parent education level, and place of birth as covariates. Average counts/min, age, and BMI (kg/m$^2$) were continuous; sex, Hispanic/Latino background, parent education, and place of birth were categorical. Hispanic/Latino background had seven levels; 0=Mexican, 1=Central American, 2=Cuban, 3=Dominican, 4=Puerto Rican, 5=South American, 6=Other or mixed. Parent education level had three levels; 0=less than high school, 1=high school or equivalent, 2=more than high school. Place of birth was dichotomous where 0=non-US born and 1=US-born. Sex = female, Hispanic/Latino background = Mexican, parent education = less than high school, and place of birth = non-US born were the reference levels. Age was centered at 12 years to coincide with the average age of the study sample. The SURVEYREG procedure in SAS was used to account for the HCHS/SOL Youth survey design. Estimates of average counts/min that were significant at the 0.05, 0.01, and 0.001 significance levels using Type III tests of fixed effects were identified by superscripts 1,2, or 3, respectively, in Table 4.5.

## 4.3 Results

The majority of the SOL Youth sample (82%) participated in the accelerometer portion of the study (Table 4.1). Approximately 40% of participants were female and 44% were between 8 and 12 years old. Thirty-nine percent of participants were classified as overweight, obese, or severely obese. There was no gain in using $m = 10$ versus $m = 5$ imputations based on relative efficiency (0.99 vs. 0.98) thus only results for $m = 5$ imputations are presented. A handful of covariates used in the logistic regression model to predict the probability of being adherent for the IPW approach had a negligible amount of missing data (Table 4.2). However, MI was performed to impute the covariates with missing values in order to increase the precision of the predicted probabilities yielded by the logistic regression model. The MGMM imputation approach for $\gamma = 1$, produced the lowest mean estimates of average counts/min and generally the lowest estimates of standard deviation compared to the other five approaches (Table 4.3). Among all of the approaches, the MGMM approach with $\gamma = 1$ and $\gamma \neq 1$ yielded the lowest estimates of standard deviation (Table 4.3). In general, MI produced the highest estimates of average counts/min compared to the other approaches. The mean and standard deviation estimates yielded by the IPW and the adherent approaches were the most identical among the approaches. This was most likely due to the underlying analysis samples being identical (Table 4.3). There were no extreme differences in point estimates of average counts/min between the approaches when missing data are imputed at the interval, day, or participant levels.

The log-linear survey regression results were similar for approaches 1 (Actical), 2 (Adherent), 3 (IPW), and 5 (MGMM) (Table 4.4). The survey regression results based on the MI data were on a different scale compared to the other approaches as average counts/min were assumed to be continuous (Table 4.6). The model coefficient and standard error estimates were smallest for both MGMM approaches with the exception of the sex variable Male (Table 4.4). Age was negatively associated with average counts/min for both MGMM approaches and positively associated with average counts/min for the remaining approaches. Sex was positively associated with average counts/min for both MGMM approaches and negatively associated with the outcome for the remaining approaches (Table 4.4). Both MGMM approaches produced smaller standard error estimates, except for sex, compared to the other approaches. The MGMM approach with $\gamma \neq 1$ yielded smaller standard error estimates compared to the approach with $\gamma = 1$.

After adjusting for age, sex, Hispanic/Latino background, parent education, and place of birth, average counts/min had a very small effect, range [-0.012, 0.007], on BMI (Table 4.5). In general, a positive model parameter estimate for average counts/min would be interpreted as "Average counts/min was positively associated with BMI (e.g., Adherent and IPW approaches)". Conversely, a negative model parameter estimate for average counts/min would be interpreted as "Average counts/min was negatively associated with BMI (e.g., Actical, Lee, MI, and MGMM)". The significant p-values for the positive model parameter estimates for average counts/min were consistent with findings in the literature as well as the U.S. Department of Health and Human Services (Table 4.5) with regard to the relationship between physical activity and BMI.[14,35,62] In other words, an increase in average counts/min results in a lower average BMI. The conclusion of whether or not average counts/min is associated with BMI is consistent for the Actical, MI, Lee, and MGMM approaches. A statistically significant association with BMI was observed for the Actical (p-value < 0.05), MI (p-value < 0.01), Lee (p-value < 0.01), and both MGMM approaches (p-value < 0.001) (Table 4.5). The results from the Adherent approach were only marginally non-significant, p-value = 0.0503. The largest p-value was observed for the IPW approach, p-value = 0.4514.

## 4.4 Discussion

Obtaining accurate measurements of physical activity and sedentary behavior from accelerometers is vital to understanding how these variables affect health outcomes. While there are many advantages to using accelerometer data in epidemiologic studies, analysis of the data poses a number of challenges. Missing accelerometer data is an inherent issue in free-living accelerometer studies. In Chapters 2-3, the multilevel generalized mixed model was presented as statistical approach to impute missing accelerometer data that incorporates the use of all available data, variability from within and between participants, and is well suited to account for multivariate count data under a complex survey design. In these previous chapters, we learned that the MGMM approach is useful for gaining insight into how average counts/min may be accumulated during accelerometer wear versus nonwear periods. For example, the rate of counts/min may be lower during nonwear versus wear segments within an interval. The results from the simulation study to assess the performance of the multilevel generalized mixed model were inconclusive and thus, did not provide evidence in support of using the method to impute missing accelerometry. In this chapter, the goal was to compare the MGMM imputation approach with five different approaches used for handling missing accelerometer data using data from the HCHS/SOL Youth Study. Moreover, we investigated whether or not imputing data at the interval, day, or participant level made a difference on point estimates and tests for associations between average counts/min and BMI.

No extreme differences were observed in average counts/min among any of approaches presented. After imputing data at the interval level with the MGMM approaches and at the day level with Lee's approach and then aggregating both at the participant level; the estimates of average counts/min were similar to what was produced by the MI approach. The standard deviations for the average counts/min estimates were smallest for both MGMM approaches (i.e., $\gamma = 1$ and $\gamma \neq 1$). Based on these results and depending on the research question, the potential payoff of imputing missing data at the interval or day level may not warrant the extra work required. The same conclusion could be made based on the results from the log-linear survey regression analysis (Table 4.4).

After adjusting for age, sex, Hispanic/Latino background, parent education level, and place of birth, the difference in predicted value of BMI for a 10-unit or 20-unit increase in average counts/minute is small and statistically significant (Table 4.5). In the context of analyzing the association between average counts/min and BMI, due to the conflicting directions of associations and statistical significance of associated p-values, imputing missing accelerometer data at the interval or day level is suggested. The p-values for the association between average counts/min and BMI were the smallest for the MGMM approaches; they provided the strongest evidence in support of an association. The direction of the association for the MI, Lee, and MGMM imputation approaches were consistent with the scientific literature; higher levels of physical activity are associated with lower BMI. The model employing the IPW showed a small but positive association between average counts/min and BMI, which conflicts with the literature.

Treating average counts/min as continuous raises one concern. Count data has traditionally been analyzed using linear regression due to the simplicity of the procedure.[7] One issue with this approach is that linear regression can produce non-integer and negative predicted counts. These values are not possible from the Actical output. Zero counts/min (per interval) was the most common value observed in the SOL Youth data leading to a distribution that was heaviest at the zero boundary (i.e., right skewed). For future studies, we recommend applying an approach that is better equipped to handle the characteristics of different types of accelerometer data (i.e., counts, raw signal, etc.) instead of reaching for the ad hoc approach of linear regression. Poisson, Negative Binomial, and Zero-inflated regression are approaches with greater model flexibility compared to linear models.

There were a couple of limitations to this study. First, the inconclusive results of the simulation study to assess the performance of the MGMM method (Chapter 3) raise some concern around the feasibility and reliability of the technique. Additional simulations studies should be performed to determine if the increased precision in parameter estimates warrants the amount of work required to implement the MGMM technique for imputation. Second, BMI for the youth study population was positively skewed with a few outliers. Model diagnostics indicated that a linear model may not be appropriate for the data. It is more common to analyze age- and sex-specific BMI percentiles such as those based on the U.S. Centers for Disease Control and Prevention (CDC) growth reference curves.[60,63] In youth, the absolute BMI is typically not utilized as a marker of risk because the measures which constitute BMI (weight and height) change as a function of normal growth and maturation.[63]

Due to the technological advancements of accelerometers in recent years, there is a shift from analyzing count data to analyzing raw signal data.[7] In light of this evolution, researchers will need to broaden their approaches to characterizing accelerometer data.[7] Analyzing raw signal data from accelerometers will help researchers better classify different types of physical activity and sedentary behavior, however, missing data due to nonwear will be an ongoing issue. Based on results from this study, imputation of missing data at the participant level may be sufficient for describing physical activity using means, medians, standard deviations, and other similar point estimates. However, if the research goal is to investigate the association between average counts/min and a health outcome, imputing missing values at the smallest unit possible (e.g., hour, interval, day, etc.), and then aggregating at the participant level, may reduce the potential for making a type 2 error (i.e., failing to identify an association when there actually is one).

## 4.5 Tables and Figures

Table 4.1. Weighted descriptive characteristics for complete SOL Youth sample and the Actical sample, SOL Youth Study 2012-2014

| | SOL Youth Sample | | SOL Youth Actical Sample | |
|---|---|---|---|---|
| Characteristic | Total n | Weighted Percent | Total n | Weighted Percent |
| Overall | 1,466 | 100.0 | 1,238 | 82.0 |
| Gender: | | | | |
| Female | 738 | 48.8 | 631 | 40.4 |
| Male | 728 | 51.2 | 607 | 41.6 |
| Age (yr): | | | | |
| 8-12 | 844 | 54.2 | 710 | 43.9 |
| 13-14 | 371 | 22.3 | 310 | 18.1 |
| 15-16 | 251 | 23.6 | 218 | 19.9 |
| BMI (kg/m$^2$): | | | | |
| Underweight | 37 | 2.8 | 31 | 2.2 |
| Normal | 712 | 50.4 | 594 | 40.8 |
| Overweight | 305 | 20.1 | 267 | 16.9 |
| Obese | 262 | 16.8 | 212 | 13.4 |
| Severely obese | 150 | 9.9 | 134 | 8.7 |

Table 4.2. Covariates included in the logistic regression model to predict the probability that a participant is adherent.

| | | | Missing data | |
|---|---|---|---|---|
| Covariate | Description | Scale | N | % |
| GENDER | Gender | Binary | 0 | 0 |
| AGE | Age | Continuous | 0 | 0 |
| BMI (kg/m$^2$) | Body mass index | Continuous | 0 | 0 |
| SITE | 4-level classification of center | Nominal | 0 | 0 |
| BACKGROUND_C7 | 7-level Hispanic/Latino background | Nominal | 25 | 1.7 |
| YOUTH_WEIGHT_NORM_OVERALL | Sampling weight | Continuous | 0 | 0 |
| PAE_TRANSPORT_WK | Transportation activity times/week | Continuous | 2 | 0.1 |
| PAE_SPORT_WK | Sports/exercise activity times/week | Continuous | 3 | 0.2 |
| PAE_SCHOOL_WK | School activity times/week | Continuous | 3 | 0.2 |
| $VO_2$ MAX (ml/kg/min) | Measure of aerobic fitness based on step test | Continuous | 59 | 4.0 |
| STRAT | Stratification | Nominal | 0 | 0 |
| IMGEN | 3-level immigrant generation | Nominal | 40 | 2.7 |
| LANG_PREF | Language preference | Binary | 4 | 0.3 |

Table 4.3. Mean (SD) average total counts/minute per person based on CCA (Actical), adherent participants, MGMM and Lee imputation approaches, and IPW.

| Characteristic | Actical N=1,238, no imputation | Adherent (≥ 3 da w/ ≥ 8 h/d) N=1,104, no imputation | IPW N=1,104 | MI N = 1,466 | Lee N=1,238 | MGMM $\gamma \neq 1$ N=1,238 | MGMM $\gamma = 1$ N=1,238 |
|---|---|---|---|---|---|---|---|
| Overall | 223.8 (5.6) | 232.0 (5.5) | 230.7 (5.6) | 238.9 (4.7) | 236.4 (4.7) | 235.7 (3.7) | 220.0 (3.6) |
| Sex | | | | | | | |
| Male | 243.9 (8.2) | 249.1 (8.4) | 245.9 (8.5) | 255.9 (7.2) | 255.6 (6.7) | 253.6 (5.2) | 237.8 (5.0) |
| Female | 203.1 (6.8) | 214.9 (7.1) | 215.3 (7.3) | 220.9 (6.0) | 216.7 (5.9) | 217.4 (4.7) | 202.1 (4.8) |
| Age (yr) | | | | | | | |
| 8-12 | 254.9 (6.8) | 262.0 (7.2) | 261.4 (7.2) | 267.5 (6.2) | 266.8 (6.1) | 266.4 (4.8) | 250.5 (4.8) |
| 13-14 | 192.3 (8.5) | 199.4 (8.1) | 200.6 (8.6) | 213.2 (7.4) | 205.5 (6.5) | 207.3 (5.2) | 191.8 (5.1) |
| 15-16 | 183.7 (13.2) | 187.8 (11.2) | 187.3 (11.4) | 197.3 (9.5) | 197.4 (9.0) | 194.0 (6.6) | 179.3 (6.6) |
| BMI (kg/m$^2$): | | | | | | | |
| Underweight | 239.8 (32.1) | 255.1 (32.3) | 254.8 (31.6) | 257.9 (27.3) | 254.0 (28.7) | 264.2 (23.9) | 250.3 (22.9) |
| Normal | 223.8 (6.7) | 235.4 (6.8) | 233.4 (7.1) | 244.1 (6.0) | 242.9 (6.0) | 242.6 (4.6) | 226.8 (4.7) |
| Overweight | 234.3 (13.2) | 243.5 (15.1) | 243.8 (14.6) | 245.5 (12.6) | 239.8 (10.7) | 237.9 (9.0) | 223.0 (8.6) |
| Obese | 224.4 (14.4) | 215.5 (8.5) | 211.3 (8.5) | 223.2 (8.0) | 227.2 (8.7) | 225.4 (6.3) | 209.8 (6.0) |
| Severely obese | 198.0 (16.4) | 211.2 (18.2) | 216.1 (19.5) | 219.9 (15.4) | 208.7 (13.9) | 207.8 (10.7) | 192.0 (10.6) |

Table 4.4. Log-linear survey regression model parameter estimates (SE) for average total counts per person using data based on the CCA, adherent participants, MGMM and Lee imputation approaches, and IPW.

| Parameter | Actical N=1,238, no imputation | Adherent (≥ 3 da w/ ≥ 8 h/d) N=1,104, no imputation | IPW N=1,104 | Lee N=1,466 | MGMM $\gamma \neq 1$ N=1,238 | MGMM $\gamma = 1$ N= 1,238 |
|---|---|---|---|---|---|---|
| Intercept | 13.0386 (0.110) | 13.0036 (0.099) | 12.9985 (0.098) | 13.0211 (0.090) | 11.5625 (0.066) | 11.5427 (0.069) |
| Age (yr) | 0.1563 (0.045) | 0.15038 (0.044) | 0.14551 (0.043) | 0.17820 (0.035) | -0.0666 (0.008) | -0.0702 (0.009) |
| Male | -0.0682 (0.011) | -0.0564 (0.010) | -0.0570 (0.010) | -0.0558 (0.008) | 0.21540 (0.039) | 0.2287 (0.042) |
| BMI (kg/m$^2$) | -0.0141 (0.003) | -0.0151 (0.003) | -0.0147 (0.003) | -0.011 (0.002) | -0.0076 (0.003) | -0.0081 (0.003) |

Table 4.5. Adjusted linear regression coefficients and standard errors (SE) for the association between average counts/min per participant and BMI ($kg/m^2$).

| Parameter | Actical | Adherent | IPW | MI | Lee | MGMM $\gamma \neq 1$ | MGMM $\gamma = 1$ |
|---|---|---|---|---|---|---|---|
| Intercept | 23.115 (0.774) | 21.094 (1.604) | 26.526 (1.484) | 23.683 (0.682) | 23.865 (0.810) | 25.114 (1.011) | 24.942 (0.969) |
| Average counts/min | -0.003 (0.002)[1] | 0.007 (0.004) | 0.003 (0.004) | -0.005 (0.002)[2] | -0.006 (0.002)[2] | -0.011 (0.003)[3] | -0.012 (0.003)[3] |
| (Age-12) (yr) | 0.769 (0.083) | 1.170 (0.241) | 0.656 (0.139) | 0.684 (0.071) | 0.725 (0.082) | 0.641 (0.087) | 0.639 (0.086) |
| Male | -0.321 (0.441) | -2.819 (1.627) | -2.077 (1.222) | -0.319 (0.402) | -0.214 (0.440) | -0.027 (0.440) | -0.023 (0.443) |
| Central American | -0.971 (0.964) | -0.390 (3.077) | -4.121 (2.238) | -0.862 (0.829) | -1.093 (0.965) | -1.300 (0.974) | -1.244 (0.965) |
| Cuban | -0.989 (0.809) | -1.873 (2.337) | -5.004 (3.766) | -0.849 (0.729) | -1.134 (0.809) | -1.369 (0.805) | -1.336 (0.805) |
| Dominican | 1.035 (0.748) | 1.322 (1.572) | 2.160 (2.192) | 1.038 (0.675) | 0.942 (0.737) | 0.765 (0.729) | 0.906 (0.722) |
| Puerto Rican | 1.018 (0.795) | -0.362 (2.702) | -3.987 (1.476) | 0.785 (0.709) | 1.002 (0.806) | 0.824 (0.814) | 0.937 (0.809) |
| South American | -0.030 (0.944) | -1.836 (2.565) | -2.311 (1.409) | -0.340 (0.806) | -0.078 (0.926) | -0.145 (0.945) | -0.105 (0.954) |
| Other or mixed | 0.861 (0.802) | 1.624 (3.268) | -0.179 (2.573) | 0.400 (0.670) | 0.821 (0.791) | 0.697 (0.782) | 0.809 (0.779) |
| Parent education: High school or equivalent | -0.162 (0.552) | 3.012 (1.828) | 0.491 (1.864) | -0.341 (0.488) | -0.149 (0.544) | -0.092 (0.538) | -0.123 (0.537) |
| Parent education: More than high school | -0.752 (0.530) | 0.469 (1.414) | -3.299 (0.878) | -0.942 (0.486) | -0.815 (0.528) | -0.791 (0.526) | -0.815 (0.524) |
| Place of birth: US born | 0.254 (0.563) | 1.327 (1.566) | -1.245 (0.362) | 0.021 (0.502) | 0.222 (0.554) | 0.218 (0.551) | 0.236 (0.553) |

[1]p-value < 0.05, [2]p-value < 0.01, [3]p-value < 0.001. Survey design strata, primary sampling units, survey and IPW weights (IPW only) were included in the analysis models to account for the complex survey design. Reference levels are Female, Mexican, Parent education: Less than high school, and Place of birth: Non-US born.

Table 4.6. Survey regression (PROC SURVEYREG) model parameter estimates (SE) for average total counts/min per person (continuous) using data based on multiple imputation using the fully conditional specification. N = 1,466.

| Characteristic | Average total counts/min |
|---|---|
| Intercept | 429.9 (25.6) |
| Age (yr) | -12.8 (2.0) |
| Male | 34.2 (9.3) |
| BMI ($kg/m^2$) | -2.4 (0.7) |

## CHAPTER 5: CONCLUDING REMARKS

This dissertation used the multilevel generalized mixed model to impute missing accelerometer data. The methodology and application of the technique was demonstrated using data from The Hispanic Community Health Study/Study of Latinos, 2008 – 2011, and The Hispanic Community Health Study/Study of Latino Youth (SOL Youth), 2012 – 2014. A challenge in many free-living accelerometer studies where the data is only available through a proprietary algorithm (e.g., Actical) is nonwear. Nonwear time, resulting in a period of consecutive zero counts, later becomes missing data in the ad hoc approach, which can bias assessments of physical activity.[21] Estimates of physical activity are biased downward as counts are not recorded during nonwear.[9] To circumvent this potential bias, researchers sometimes analyze data only for participants with a minimum number of adherent days, defined as having a sufficient amount of wear time in a given day (ad hoc approach).[9,12] Non-adherent days are labelled as "missing". The concern with this approach is that it assumes physical activity is missing completely at random (MCAR) during nonwear. Based on the results from chapter two, we concluded that the rate of counts/min may not be the same for wear and nonwear periods and thus the assumption of MCAR data was not tenable.

Generalized linear mixed models are appropriate for relating changes in the mean of a discrete response variable (e.g., counts per interval) to covariates.[23] The mixed model (i.e., fixed and random effects) framework was selected as it is ideal for modeling correlated, unbalanced, and hierarchically structured data. The addition of nested random effects to these models increased the precision with which regression parameters can be estimated.[23] The results from the simulation study (Chapter 3) to assess the performance of the imputation technique were inconclusive. We expected to observe smaller estimates of percent relative bias for accurately-specified imputation models and larger estimates for misspecified models. However, we only observed this behavior for the Poisson data generation model with $\gamma = 1.3$ and Poisson evaluation model with $\gamma \neq 1.0$. In a secondary analysis, the Poisson marginal model (e.g., without random effects) for evaluation yielded estimates of percent relative bias that were approximately 10% lower than the corresponding percent relative biases produced by all of the multilevel mixed evaluation models. Thus, we concluded that the Poisson marginal mixed model was the "best" model. In addition to the conflicting results, the percent relative biases for all evaluation models were appreciably greater than 10%, which was unsatisfactory. When applied to a different study population (i.e., youth), the multilevel mixed model for imputation worked well (Chapter 4).

Based on result from Chapter 4, imputation of missing data at the participant level for youth may be sufficient for describing physical activity data using the mean, median, standard deviation, or similar point estimates. However, if the research goal is to investigate the association between average counts/min and a health outcome of interest, imputing missing values at the smallest unit possible (e.g., hour, interval, day) and then aggregating at the participant level may reduce the potential for making a type 2 error (i.e., failing to identify an association when there actually is one).

The field of accelerometry-based research is expanding. As the accelerometry technology evolves, investigators are increasingly interested in analyzing the data in its purer forms such as the raw signals. In light of this evolution, researchers will need to broaden their approaches to characterizing accelerometer data.[7] Analyzing raw signal data from accelerometers will help researchers better classify different types of physical activity and sedentary behavior, however, missing data due to nonwear will be an ongoing issue. With regard to the public health impact of missing accelerometry; depending on the variable of interest, inaccurately describing a particular population based on biased physical activity and sedentary behavior estimates could lead to misinformed and detrimental decision making. For example, if an intervention will only be delivered to a certain group of individuals who do not meet physical activity guidelines, mis-specifying the assignment as to whether or not guidelines were met would mean that some individuals would miss the intervention when they actually needed it. For researchers analyzing accelerometry, as an initial step, we recommend exploring the data distribution, proportion and pattern of missing data, as well as the missing mechanism. This type of exploratory analysis will help determine how data analysis should proceed. Once an understanding of the data has been established, we recommend applying analytical approaches that are better equipped to handle the characteristics of different types of accelerometer data (i.e., counts, raw signal, etc.). For example, Poisson, Negative Binomial, and Zero-inflated regression models have greater model-fitting flexibility compared to linear models.

One proposal for future study is to use a statistical model (i.e., Zero-Inflated Poisson and Zero-Inflated Negative Binomial) to identify true zero counts per epoch associated with sedentary behavior versus zero counts per epoch associated with accelerometer nonwear, which are then defined as missing data. The innovation of this proposed study is twofold: (1) No assumptions about wear and nonwear intervals would be made. This differs from the conventional approach of classifying intervals of consecutive zero counts/epoch (e.g., 20, 30, 60, 90 minutes) as nonwear.[9,14,21,22,24,48] (2) Each zero count per epoch would be identified as either nonwear or sedentary behavior.

The significance of this proposed study is that identifying nonwear versus sedentary behavior zeros at the epoch level versus some higher level (i.e., hour or interval) may yield more accurate estimates of sedentary behavior and nonwear. This is to say; traditional methods may "inaccurately" classify certain zeros as nonwear which may lead to an underestimation of wear time and sedentary behavior.

Overall, we hypothesized that (1) accelerometer average counts/min are higher for wear versus nonwear segments in an interval, (2) percent relative bias would be smaller for multilevel generalized mixed imputation evaluation models that were concordant with multilevel generalized mixed data generation models, and that (3) there would be a significant association between average count/min and BMI when missing values were imputed at the interval level using the multilevel generalized mixed model. We found that accelerometer average counts/min were higher for wear versus nonwear segments in an interval and concluded that the MCAR assumption of the ad hoc approach was not tenable. We did not find any meaningful evidence in support of percent relative biases being smaller for multilevel generalized mixed imputation evaluation models that were concordant with multilevel generalized mixed data generation models. Thus, based on the simulation results alone, we would not recommend the multilevel generalized mixed model method for imputation. However, we recommend further simulations studies be conducted to better asses the multilevel generalized mixed model method for imputation. Lastly, we found evidence in support of there being an association between average count/min and BMI when missing values were imputed at the interval level using the multilevel generalized mixed model. Thus, we concluded that imputing missing values at the smallest unit possible (e.g., interval), and then aggregating at the participant level, may reduce the potential for making a type 2 error (i.e., failing to identify an association when there actually is one).

## REFERENCES


1. *Physical Activity Guidelines; Chapter 4: Active Adults*. Department of Health & Human Services; 2008. https://health.gov/paguidelines/guidelines/chapter3.aspx. Accessed September 7, 2016.

2. *Physical Activity Guidelines; Chapter 3: Active Children and Adolescents*. Department of Health & Human Services; 2008. https://health.gov/paguidelines/guidelines/chapter3.aspx. Accessed September 7, 2016.

3. Ogden CL, Carroll MD, Kit BK, Flegal KM. Prevalence of childhood and adult obesity in the united states, 2011-2012. *JAMA*. 2014;311(8):806-814. doi:10.1001/jama.2014.732

4. *Physical Activity Guidelines; Appendix 6: Glossary of Terms*. Department of Health & Human Services; 2008. https://health.gov/paguidelines/guidelines/chapter3.aspx. Accessed July 19, 2017.

5. Troiano RP, Pettee Gabriel KK, Welk GJ, Owen N, Sternfeld B. Reported physical activity and sedentary behavior: why do you ask? *J Phys Act Health*. 2012;9 Suppl 1:S68-75.

6. Lee I-M, Shiroma EJ. Using accelerometers to measure physical activity in large-scale epidemiological studies: issues and challenges. *Br J Sports Med*. 2014;48(3):197-201. doi:10.1136/bjsports-2013-093154

7. Troiano RP, McClain JJ, Brychta RJ, Chen KY. Evolution of accelerometer methods for physical activity research. *Br J Sports Med*. 2014;48(13):1019-1023. doi:10.1136/bjsports-2014-093546

8. Strath SJ, Bassett DR, Swartz AM. Comparison of the college alumnus questionnaire physical activity index with objective monitoring. *Ann Epidemiol*. 2004;14(6):409-415. doi:10.1016/j.annepidem.2003.07.001

9. Catellier DJ, Hannan PJ, Murray DM, et al. Imputation of Missing Data When Measuring Physical Activity by Accelerometry: *Med Sci Sports Exerc*. 2005;37(Supplement):S555-S562. doi:10.1249/01.mss.0000185651.59486.4e

10. John D, Freedson P. ActiGraph and Actical Physical Activity Monitors: A Peek under the Hood. *Med Sci Sports Exerc*. 2012;44:S86-S89. doi:10.1249/MSS.0b013e3182399f5e

11. Evenson KR, Wen F. Performance of the ActiGraph accelerometer using a national population-based sample of youth and adults. *BMC Res Notes*. 2015;8(1):7. doi:10.1186/s13104-014-0970-2

12. Mâsse LC, Fuemmeler BF, Anderson CB, et al. Accelerometer Data Reduction: A Comparison of Four Reduction Algorithms on Select Outcome Variables: *Med Sci Sports Exerc*. 2005;37(Supplement):S544-S554. doi:10.1249/01.mss.0000185674.09066.8a

13. Heil DP, Brage S, Rothney MP. Modeling Physical Activity Outcomes from Wearable Monitors: *Med Sci Sports Exerc*. 2012;44:S50-S60. doi:10.1249/MSS.0b013e3182399dcc

14. Troiano RP, Berrigan D, Dodd KW, MâSse LC, Tilert T, Mcdowell M. Physical Activity in the United States Measured by Accelerometer: *Med Sci Sports Exerc*. 2008;40(1):181-188. doi:10.1249/mss.0b013e31815a51b3

15. Dong Y, Peng C-YJ. Principled missing data methods for researchers. *SpringerPlus*. 2013;2(1):222. doi:10.1186/2193-1801-2-222

16. Schafer JL. Multiple imputation: a primer. *Stat Methods Med Res*. 1999;8(1):3-15. doi:10.1177/096228029900800102

17. Allison PD. Missing Data. In: *The SAGE Handbook of Quantitative Methods in Psychology*. 1 Oliver's Yard, 55 City Road, London EC1Y 1SP United Kingdom: SAGE Publications Ltd; 2009:72-90. http://methods.sagepub.com/book/sage-hdbk-quantitative-methods-in-psychology/n4.xml. Accessed January 23, 2017.

18. Liu B, Yu M, Graubard BI, Troiano RP, Schenker N. Multiple imputation of completely missing repeated measures data within person from a complex sample: application to accelerometer data in the National Health and Nutrition Examination Survey: Multiple imputation of the accelerometer data in the NHANES. *Stat Med*. 2016;35(28):5170-5188. doi:10.1002/sim.7049

19. Lee PH. Data imputation for accelerometer-measured physical activity: the combined approach. *Am J Clin Nutr*. 2013;97(5):965-971. doi:10.3945/ajcn.112.052738

20. Boyer WR, Wolff-Hughes DL, Bassett DR, Churilla JR, Fitzhugh EC. Accelerometer-Derived Total Activity Counts, Bouted Minutes of Moderate to Vigorous Activity, and Insulin Resistance: NHANES 2003–2006. *Prev Chronic Dis*. 2016;13. doi:10.5888/pcd13.160159

21. Xu SY, Nelson S, Kerr J, et al. Statistical approaches to account for missing values in accelerometer data: Applications to modeling physical activity. *Stat Methods Med Res*. July 2016:096228021665711. doi:10.1177/0962280216657119

22. Lee JA, Gill J. Missing value imputation for physical activity data measured by accelerometer. *Stat Methods Med Res*. March 2016:096228021663324. doi:10.1177/0962280216633248

23. Fitzmaurice GM, Laird NM, Ware JH. *Applied Longitudinal Analysis*. 2nd ed. Hoboken, N.J: Wiley; 2011.

24. Morris JS, Carroll RJ. Wavelet-based functional mixed models. *J R Stat Soc Ser B Stat Methodol*. 2006;68(2):179-199. doi:10.1111/j.1467-9868.2006.00539.x

25. Johnstone IM. Wavelets and the theory of non-parametric function estimation. *Philos Trans R Soc Math Phys Eng Sci*. 1999;357(1760):2475-2493. doi:10.1098/rsta.1999.0443

26. Lee KJ, Carlin JB. Multiple Imputation for Missing Data: Fully Conditional Specification Versus Multivariate Normal Imputation. *Am J Epidemiol*. 2010;171(5):624-632. doi:10.1093/aje/kwp425

27. Berglund P A. Multiple Imputation Using the Fully Conditional Specification Method: A Comparison of SAS®, Stata, IVEware, and R. In: *Proceedings of the SAS® Global Forum 2015 Conference*. Cary, NC: SAS Institute Inc.; 2015. http://support.sas.com/resources/papers/proceedings10/265-2010.pdf. Accessed July 24, 2016.

28. Soley-Bori, Marina. *Dealing with Missing Data: Key Assumptions and Methods for Applied Analysis*. Boston University; 2013. https://www.bu.edu/sph/files/2014/05/Marina-tech-report.pdf. Accessed February 18, 2018.

29. Rubin DB. *Multiple Imputation for Nonresponse in Surveys*. New York: Wiley; 1987.

30. Ibrahim JG, Molenberghs G. Missing data methods in longitudinal studies: a review. *TEST*. 2009;18(1):1-43. doi:10.1007/s11749-009-0138-x

31. Mallinckrodt C. *Preventing and Treating Missing Data in Longitudinal Clinical Trials a Practical Guide*. Cambridge: Cambridge University Press; 2013.

32. Donner A. The Relative Effectiveness of Procedures Commonly Used in Multiple Regression Analysis for Dealing with Missing Values. *Am Stat*. 1982;36(4):378. doi:10.2307/2683092

33. Little RJA, Rubin DB. *Statistical Analysis with Missing Data*. 2nd ed. Hoboken, N.J: Wiley; 2002.

34. Seaman SR, White IR. Review of inverse probability weighting for dealing with missing data. *Stat Methods Med Res*. 2013;22(3):278-295. doi:10.1177/0962280210395740

35. 2018 Physical Activity Guidelines Advisory Committee. *2018 Physical Activity Guidelines Advisory Committee Scientific Report*. Washington, DC: Department of Health & Human Services; 2018. https://health.gov/paguidelines/second-edition/report/pdf/PAG_Advisory_Committee_Report.pdf. Accessed March 3, 2018.

36. *Facts about Physical Activity*. Department of Health & Human Services; 2014. https://www.cdc.gov/physicalactivity/data/facts.htm. Accessed September 26, 2016.

37. Trost SG, Pate RR, Freedson PS, Sallis JF. Using objective physical activity measures with youth: How many days of monitoring are needed? *Med Sci Sports Exerc*. 2000;32(2):426-431.

38. Zhu W, Owen N, eds. *Sedentary Behavior and Health: Concepts, Assessments, and Interventions*. Champaign, IL: Human Kinetics; 2017.

39. Wong SL, Colley R, Connor Gorber S, Tremblay M. Actical accelerometer sedentary activity thresholds for adults. *J Phys Act Health*. 2011;8(4):587-591.

40. Chinapaw MJM, de Niet M, Verloigne M, De Bourdeaudhuij I, Brug J, Altenburg TM. From Sedentary Time to Sedentary Patterns: Accelerometer Data Reduction Decisions in Youth. Courvoisier DS, ed. *PLoS ONE*. 2014;9(11):e111205. doi:10.1371/journal.pone.0111205

41. Evenson KR, Herring AH, Wen F. Accelerometry-Assessed Latent Class Patterns of Physical Activity and Sedentary Behavior With Mortality. *Am J Prev Med*. 2017;52(2):135-143. doi:10.1016/j.amepre.2016.10.033

42. Healy GN, Dunstan DW, Salmon J, et al. Breaks in Sedentary Time: Beneficial associations with metabolic risk. *Diabetes Care*. 2008;31(4):661-666. doi:10.2337/dc07-2046

43. Altenburg TM, Rotteveel J, Dunstan DW, Salmon J, Chinapaw MJM. The effect of interrupting prolonged sitting time with short, hourly, moderate-intensity cycling bouts on cardiometabolic risk factors in healthy, young adults. *J Appl Physiol*. 2013;115(12):1751-1756. doi:10.1152/japplphysiol.00662.2013

44. Peddie MC, Bone JL, Rehrer NJ, Skeaff CM, Gray AR, Perry TL. Breaking prolonged sitting reduces postprandial glycemia in healthy, normal-weight adults: a randomized crossover trial. *Am J Clin Nutr*. 2013;98(2):358-366. doi:10.3945/ajcn.112.051763

45. Sorlie PD, Avilés-Santa LM, Wassertheil-Smoller S, et al. Design and Implementation of the Hispanic Community Health Study/Study of Latinos. *Ann Epidemiol*. 2010;20(8):629-641. doi:10.1016/j.annepidem.2010.03.015

46. LaVange LM, Kalsbeek WD, Sorlie PD, et al. Sample Design and Cohort Selection in the Hispanic Community Health Study/Study of Latinos. *Ann Epidemiol*. 2010;20(8):642-649. doi:10.1016/j.annepidem.2010.05.006

47. Evenson KR, Sotres-Alvarez D, Deng Y, et al. Accelerometer Adherence and Performance in a Cohort Study of US Hispanic Adults: *Med Sci Sports Exerc*. 2015;47(4):725-734. doi:10.1249/MSS.0000000000000478

48. Choi L, Liu Z, Matthews CE, Buchowski MS. Validation of Accelerometer Wear and Nonwear Time Classification Algorithm: *Med Sci Sports Exerc*. 2011;43(2):357-364. doi:10.1249/MSS.0b013e3181ed61a3

49. Merchant G, Buelna C, Castañeda SF, et al. Accelerometer-measured sedentary time among Hispanic adults: Results from the Hispanic Community Health Study/Study of Latinos (HCHS/SOL). *Prev Med Rep*. 2015;2:845-853. doi:10.1016/j.pmedr.2015.09.019

50. Thorp AA, Owen N, Neuhaus M, Dunstan DW. Sedentary Behaviors and Subsequent Health Outcomes in Adults. *Am J Prev Med*. 2011;41(2):207-215. doi:10.1016/j.amepre.2011.05.004

51. Rabe-Hesketh S, Skrondal A, Pickles A. Maximum likelihood estimation of limited and discrete dependent variable models with nested random effects. *J Econom*. 2005;128(2):301-323. doi:10.1016/j.jeconom.2004.08.017

52. Berglund P, Heeringa S. *Multiple Imputation of Missing Data Using SAS*. Cary, N.C: SAS Institute; 2014.

53. Zeger SL, Liang K-Y. Longitudinal Data Analysis for Discrete and Continuous Outcomes. *Biometrics*. 1986;42(1):121. doi:10.2307/2531248

54. Rubin DB. Inference and Missing Data. *Biometrika*. 1976;63(3):581. doi:10.2307/2335739

55. Carle AC. Fitting multilevel models in complex survey data with design weights: Recommendations. *BMC Med Res Methodol*. 2009;9(1). doi:10.1186/1471-2288-9-49

56. Graham JW, Olchowski AE, Gilreath TD. How many imputations are really needed? Some practical clarifications of multiple imputation theory. *Prev Sci Off J Soc Prev Res*. 2007;8(3):206-213. doi:10.1007/s11121-007-0070-9

57. Mistler, Stephen A. A SAS® Macro for Applying Multiple Imputation to Multilevel Data. In: *SAS Paper 438-2013*. Arizona State University: SAS Institute Inc.; 2013. https://support.sas.com/resources/papers/proceedings13/438-2013.pdf. Accessed April 11, 2018.

58. Isasi CR, Carnethon MR, Ayala GX, et al. The Hispanic Community Children's Health Study/Study of Latino Youth (SOL Youth): design, objectives, and procedures. *Ann Epidemiol*. 2014;24(1):29-35. doi:10.1016/j.annepidem.2013.08.008

59. Gallo LC, Roesch SP, McCurley JL, et al. Youth and Caregiver Physical Activity and Sedentary Time: HCHS/SOL Youth. *Am J Health Behav*. 2017;41(1):67-75. doi:10.5993/AJHB.41.1.7

60. *A SAS Program for the 2000 CDC Growth Charts (Ages 0 to <20 Years)*. Centers for Disease Control and Prevention https://www.cdc.gov/nccdphp/dnpao/growthcharts/resources/sas.htm. Accessed March 26, 2014.

61. O'Kelly M, Ratitch B. *Clinical Trials with Missing Data: A Guide for Practitioners*. Chichester, West Sussex: John Wiley & Sons Inc; 2014.

62. Lee O, Lee D, Lee S, Kim YS. Associations between Physical Activity and Obesity Defined by Waist-To-Height Ratio and Body Mass Index in the Korean Population. Tauler P, ed. *PLOS ONE*. 2016;11(7):e0158245. doi:10.1371/journal.pone.0158245

63. Harrington DM, Staiano AE, Broyles ST, Gupta AK, Katzmarzyk PT. BMI percentiles for the identification of abdominal obesity and metabolic risk in children and adolescents: evidence in support of the CDC 95th percentile. *Eur J Clin Nutr*. 2013;67(2):218-222. doi:10.1038/ejcn.2012.203